Assessment of disability in multiple sclerosis.

Sharrack, Basil

ASSESSMENT OF DISABILITY IN MULTIPLE SCLEROSIS

Basil Sharrack

A thesis submitted for the award of the degree of
Doctor of Philosophy [Ph.D.]

Guy's, Kings and St. Thomas' School of Medicine
King's College
University of London

July 1999

ABSTRACT

A review of the existing clinical outcome measures used for assessing impairment, disability, and handicap in multiple sclerosis revealed that none is satisfactory.

I conducted a prospective study to assess the psychometric properties of the five commonly used clinical rating scales for multiple sclerosis in a cohort of 64 patients, and found that none satisfied all the requirements of an ideal outcome measure although all had some desirable properties.

I have devised a new clinical disability scale, the Guy's Neurological Disability Scale, and found it to be reliable, valid, and responsive in a cohort of 50 patients. I showed that the scale could be satisfactory administered over the telephone or via a postal questionnaire.

I have assessed the correlation between pathology as measured by total lesion load on T1- and T2-weighted brain magnetic resonance images and impairment, disability, handicap, and health related quality of life outcome measures used in multiple sclerosis. This study involved the same cohort of 50 patients and used a novel semi-automated computer assisted quantitative method. There was a modest correlation between the extent of pathology and some measures of impairment, disability and health related quality of life, but not with any handicap measure.

I concluded that clinical disability rating scales should remain the gold standard for assessing the outcome of clinical trials in multiple sclerosis, and that the Guy's Neurological Disability Scale is a valuable and promising new clinical outcome measure.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AI	Ambulation Index
CAMBS	Cambridge Multiple Sclerosis Basic Score
CI	Confidence interval
EDSS	Expanded Disability Status Scale
FIM	Functional Independence Measure
FS	Functional System
GNDS	Guy's Neurological Disability Scale
GNDS-R	Revised Guy's Neurological Disability Scale
ICC	Intraclass correlation coefficient
ICD	International Classification of Diseases
ICIDH	International Classification of Impairments, Disabilities and Handicaps
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
RC	Repeatability coefficient
SD	Standard Deviation
SNRS	Scripps Neurological Rating Scale
VAS	Visual Analogue Scale
WHO	World Health Organisation

DEDICATION

To Sawsan

My beloved wife and dearest friend

ACKNOWLEDGEMENTS

May people have been involved in this project in one way or another, but I would like to acknowledge a particular debt of gratitude to Professor Richard Hughes for his advice and encouragement, for his help in the inter-rater reliability studies, and for his continuous support and friendship.

I would also like to thank Professor John Weinman and Professor Graham Dunn for their invaluable advice and guidance, Sister Sara Soudain for her help in the inter-rater reliability and validation studies, and all the academic staff in the Department of Neurology at UMDS for making my research years as enjoyable as they have been.

I would also like to acknowledge the help of all the neurologists who participated in the validity studies, and all the patients in the MS Research Clinic at Guy's Hospital for their help and kind co-operation.

I would also like to thank the following who contributed to this project: Mr. Peter Podgorski, Dr. Tim Cox, Professor Alan Thompson, Professor M Ron, Dr. G Plant, Dr. C Fowler, Dr. S Wessely, Dr. D Wade, Professor M Wiles, Dr Luke Kartsounis, Dr. Crispin Jenkinson, Dr. Dona Lamping, Dr. Charles Poser, Mr. John Simkins, and Ares-Serono, the Multiple Sclerosis Research Trust, and the MS Resource Centre for funding this research.

This project would have not been possible had I not already received a comprehensive training in clinical neurology, and I particularly acknowledge Dr Aelwyn Davies-Jones in this respect.

I would also like to take this opportunity to thank Sawsan for her understanding, support, guidance, friendship, and much more, and Noor and Sana, who grew with this project, for brightening my life.

PUBLICATIONS ARISING FROM THIS THESIS

Sharrack, B. and Hughes, R.A.C. (1996) Clinical scales for multiple sclerosis. *Journal of the Neurological Sciences* **135**, 1-9.

Sharrack, B. and Hughes, R.A.C. (1997) Reliability of distance estimation by doctors and patients: cross sectional study. *British Medical Journal* **315**, 1652-1654.

Hughes, R.A.C. and Sharrack, B. (1998) Measuring disablement and quality of life in multiple sclerosis. *European Journal of Neurology* **5**, S41-S42 (Abstract)

Sharrack, B., Hughes, R.A.C., Soudain, S. and Dunn, G. (1999) The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* **122**, 141-159.

Sharrack, B. and Hughes, R.A.C. (1999) Scale development and the Guy's Neurological Disability Scale. *Journal of Neurology* **246**, 226

Sharrack, B. and Hughes, R.A.C. (1999) The Guy's Neurological Disability Scale: a new disability scale for multiple sclerosis. *Multiple Sclerosis* **5**, 223-233

Sharrack, B., Hughes, R.A.C., Soudain, S., Colchester, A.C. and Bello, F. (1999) The correlation between magnetic resonance imaging lesion load and disablement and quality of life outcome measures in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **66**, 270 (Abstract)

"It is the mark of the educated man and a proof of his culture that in every subject he looks for only so much precision as nature permits".

(Aristotle)

Chapter 1

HUMAN DISABLEMENT AND QUALITY OF LIFE IN NEUROLOGICAL DISORDERS

1.1 Introduction

The greatest problem facing anyone devising health-related outcome measures is to decide upon a logical, coherent, and comprehensive framework of classification which can be used in assessing the relation between disease and its consequences. The lack of such an adequate classification has hindered the development of health science research and complicated the interpretation of clinical trials.

Over the last century, clinical and public health practices have transformed in response to the reduction in mortality from infectious diseases and the increase in chronic non-fatal degenerative diseases. This transformation has led to changes in the challenges posed by disease, and the measures needed to assess its consequences. Before these changes, the International Classification of Diseases (ICD) (World Health Organisation, 1993), which had existed since 1893, provided a valuable and relevant model for studying the disease process (Thuriaux, 1995). This model represented a concept of disease which has been depicted as a sequence:

cure

aetiology → *pathology* → *manifestation* →

death

This concept provided an efficient model for disorders that could be prevented or cured, but did not encompass the non-fatal consequences of chronic disorders which have been collectively described as 'disablement' (Granger and

Gresham, 1990b; Badley, 1993; Thuriaux, 1995). The conceptual model of disease-related phenomena thus needed further expansion.

1.2 The International Classification of Impairments, Disabilities and Handicaps

In the 1970's, the World Health Organisation (WHO) tried to develop a comprehensive scheme for measuring 'disablement' which would be compatible with the principles underlying the structure of the ICD. However detailed deliberations between numerous individuals and various organisations led to the realisation that a single classification scheme conforming to the taxonomic principles of the ICD would not be satisfactory. Instead, a novel model of classification was structured along three levels, the body level (impairment – 1009 items), the person level (disability – 338 items), and the society level (handicap – 72 items). This model has been depicted as a sequence:

disease (pathology) → *impairment* → *disability* → *handicap*

This classification, which was published by the WHO in 1980 as the International Classification of Impairments, Disabilities and Handicaps (ICIDH) (World Health Organisation, 1980), provided a common language for research, and a comprehensive framework for discussing and understanding the consequences of diseases.

The original model had a basic three-level hierarchical structure (Figure 1.1). Wade (1995a) argued that diseases need to be considered at four levels: pathology, impairment, disability, and handicap. These four levels form a continuum with many grey areas in between and share a diverse chain of inter-relations (Figure 1.2).

Figure 1.1 The essential components of the ICIDH model as adapted from Granger (1990b)

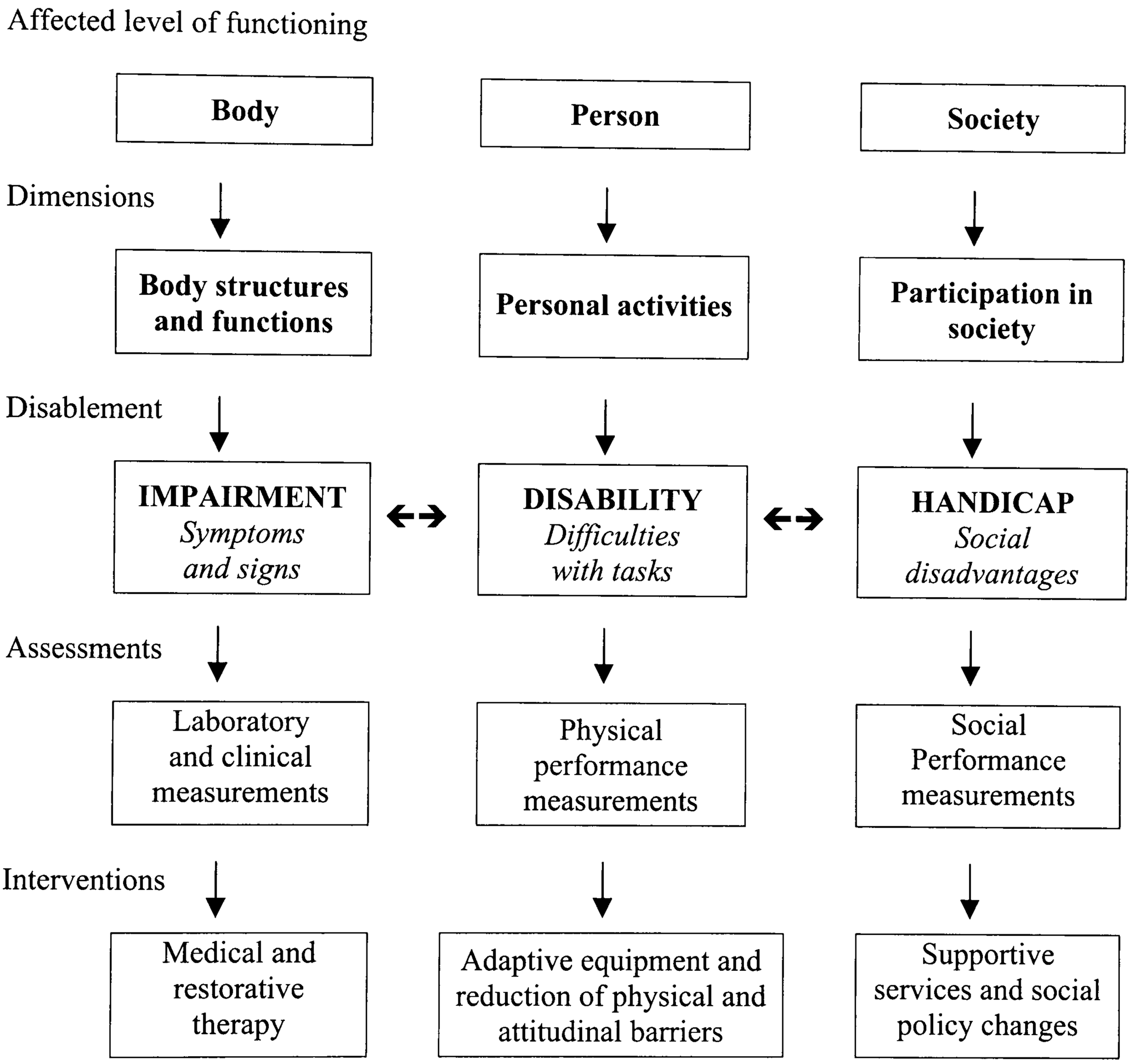
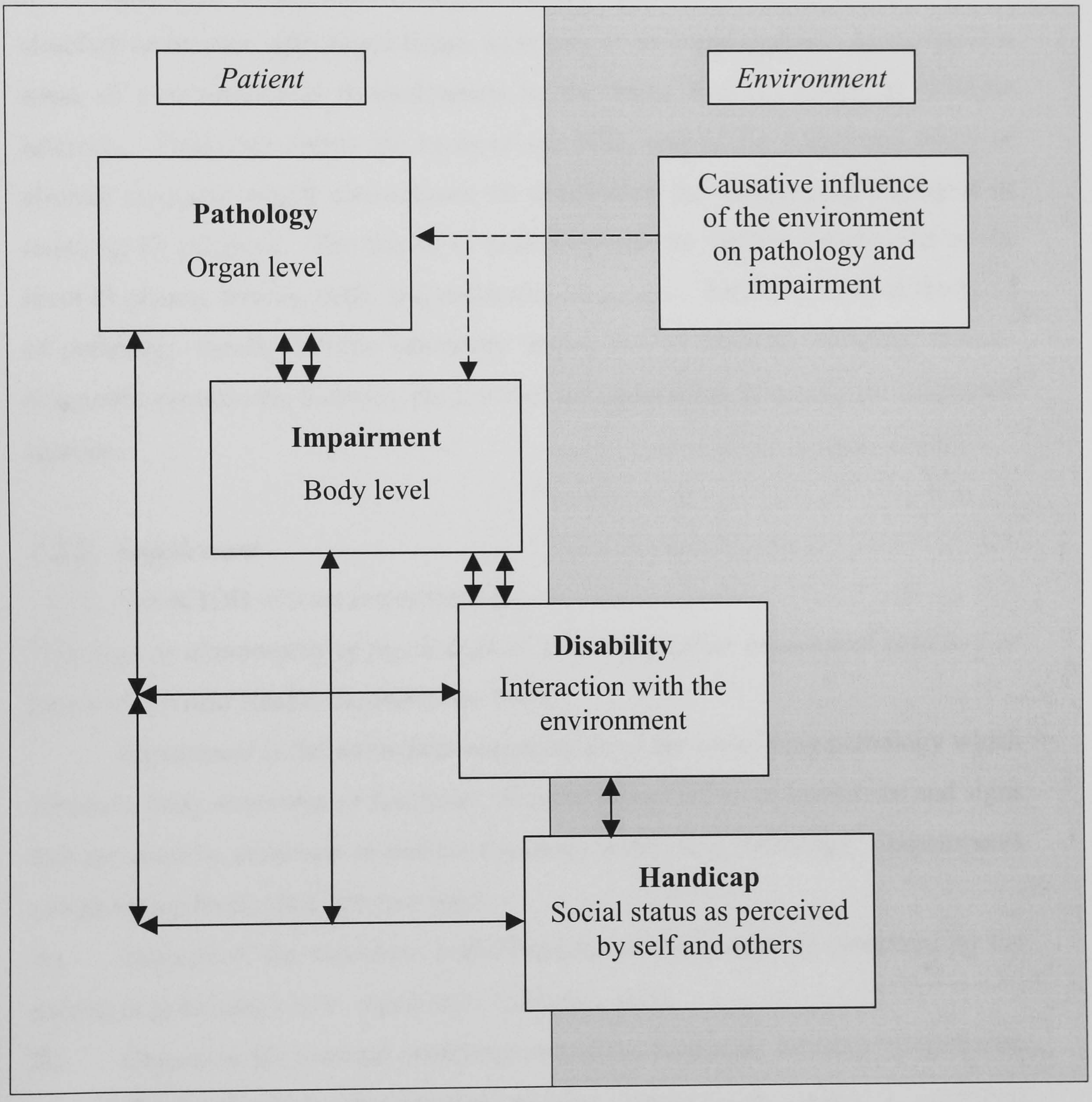


Figure 1. 2 The WHO model of disease, illustrating the relationship between the patient and the environment in the four dimensions of the ICIDH as adapted from Wade (1996). The arrows refer to the strength of these relationships



1.2.1 Pathology

The ICIDH defines the first principal event in the development of an illness as:

"A chain of causal circumstances, the 'etiology', gives rise to changes in the structural or functioning of the body, the pathology" (World Health Organisation, 1980).

Pathology is any abnormality of microscopic, macroscopic, or biochemical structure or function affecting a tissue, an organ, or an organ system. An example is areas of peri-ventricular demyelination in the brain such as occur in multiple sclerosis. Pathology forms the basis of the ICD, and is the traditional focus of clinical medicine which concentrates on diagnosing the disease and curing it or reducing its progress. The theme of pathology can be further sub-divided at the level of organs, tissues, cells, and molecular structures. Measurements at the level of pathology usually involve laboratory testing (blood analysis, imaging, electro-diagnostic procedures, biopsies, etc.), which are undertaken primarily for diagnostic reasons.

1.2.2 Impairment

The ICIDH defines impairment as:

"any loss or abnormality of psychological, physiological or anatomical structure or function" (World Health Organisation, 1980).

Impairment is the immediate consequence of the underlying pathology which relates to body structures or functions. It is the constellation of symptoms and signs that are used by clinicians to deduce the likely underlying pathology. Impairments can therefore be divided into two types:

A). Subjective: the conscious manifestations of the disease as perceived by the patient (e.g. blurred vision, numbness, weakness, etc.).

B). Objective: the external manifestations of the disease as detected by the health care professional, which often have no personal meaning to the patient (e.g. afferent pupillary defect, impaired vibration sense, brisk tendon reflexes).

Impairment describes the situation at a particular point of time, and can be temporary or permanent, intermittent or continuous, and progressive, regressive, or

static. It is often specific to a disease or a group of diseases. Consequently, many specific scales of disease severity measure impairment. Examples of impairment measures include visual acuity measures (Acheson and Sanders, 1995), the Scripps Neurological Rating Scale (Sipe et al., 1984), and thermal and vibration perception thresholds (Hughes et al., 1995).

The ICIDH model predicts a close relation between pathology and impairment. However pathology may occur without impairment and impairment may occasionally occur without pathology (Figure 1.2). Some post mortem studies, for example, have shown widespread areas of demyelination typical of multiple sclerosis in people with no recorded appropriate symptoms or appropriate clinical findings during life (Gilbert and Sadler, 1983). Conversely, impairments related to tension headache are not usually associated with any demonstrable pathology (Wade, 1996). The nature, size, and the location of the pathological process is not always directly related to the degree of impairment. Non-fluent Broca's aphasia, for example, can be a manifestation of an infarct, a tumour, or an abscess affecting the posterior part of the inferior frontal gyrus of the dominant hemisphere. Similarly, patients with a large cerebral lesion load in multiple sclerosis may remain asymptomatic, whereas others with small plaques in the brain stem or the spinal cord may have devastating symptoms and signs (Thompson et al., 1990). Finally, spastic paraparesis may result from lesions affecting the forebrain, brain stem, or cervical / dorsal spinal cord. As techniques for detecting pathology become more sensitive, the weakness in the links between pathological processes and impairment is becoming more apparent. For instance, magnetic resonance imaging in patients with clinically isolated optic neuritis often demonstrates widespread asymptomatic abnormalities in other parts of the central nervous system (Morrisey et al., 1993).

1.2.3 Disability

The ICIDH defines disability as:

"*any restriction or lack (resulting from impairment) of ability to perform an activity in the manner or within the range considered normal for a human being*" (World Health Organisation, 1980).

Disability is the functional loss which arises either as a direct consequence of the physical impairment or as a psychological response to its presence. It is characterised by functional limitation in performing customary activities or behaviours which arise at the level of the person's interaction with the immediate environment. Disabilities can be temporary or permanent, reversible or irreversible, and progressive, regressive or static. Examples include difficulties in reading, grooming, or stair climbing, or needing a stick or crutches to ambulate. Impairment is not always readily separated from disability. However as one moves from impairment to disability, the functional loss develops an increasingly more personal meaning to the patient. For example, when testing strength to determine the presence or absence of weakness, an impairment, the action of flexing the hip or the knee has no intrinsic purpose to the patient other than following the instructions given. However the same action becomes more meaningful at the level of disability when the patient experiences difficulties with walking or climbing stair. Disability is usually measured using clinical rating scales such as the Ambulation Index (Hauser et al., 1983), the Functional Independence Measure (Keith et al., 1987b), and the Barthel Index (Mahoney and Barthel, 1965).

The ICIDH model predicts that the links between pathology and the nature and extent of disability are relatively weak (Figure 1.2). Many magnetic resonance imaging studies in multiple sclerosis have shown no (Thompson et al., 1990) or only a weak relationship (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995) between total lesion load, as a measure of pathology, and clinical disability. The ICIDH model, on the other hand, predicts a relatively close relation between impairment and disability (Figure 1.2). However this relation is complicated by many intervening variables, the most important of which are the situational factors and the psychological well being of the patient. Aids, equipment, environmental adaptations and environmental changes allow patients to function more independently despite unchanging impairments. For instance, independence in some activity such as walking or stair climbing may be achieved by the provision of a foot-drop splint or moving to single storey accommodation. Patients may also learn to fulfil their behavioural goals in other ways in the presence of static impairments. For instance, patients may learn to walk

despite the presence of weak or spastic lower limbs. The relation between specific impairments and specific disabilities is therefore not static and may vary over time, as adaptation occurs, and between patients depending on their motivation, adaptability, and opportunities.

1.2.4 Handicap

The ICIDH defines handicap as:

"*a disadvantage for a given individual, resulting from an impairment or a disability, that limits or prevents the fulfilment of a role that is normal (depending on age, sex, and social and cultural factors) for that individual*" (World Health Organisation, 1980).

Handicap refers to the social consequences of impairment and disability, which arise at the level of the patient's own social roles and activities. It is reflected in the discordance between the patient's performance or status and the expectations of their social group. The distinction between disability and handicap therefore rests on the difference between performing tasks and performing roles (Bury, 1987). The most important distinguishing characteristic of handicap is that normality is judged with reference to the patient's own immediate social context (family, friends, neighbourhood, etc.), whereas normality for pathology, impairment and disability is generally judged with reference to the population at large. The evaluation of handicap is dependent on the cultural norms, so that a person may be handicapped in one group but not in another. Examples of handicap include the loss of a job or earning as a consequence of the inability to walk. Handicap is usually measured with clinical rating scales such as the Environmental Status Scale (International Federation of Multiple Sclerosis Societies, 1985), or the London Handicap Scale (Harwood et al., 1994).

The ICIDH model predicts that the relation between pathology, impairment, and handicap is weak in either direction (Figure 1.2). In practice, handicap can arise directly from impairment, or even from pathology without impairment or disability. For example, a hemianopia, an impairment, may be asymptomatic causing no disability to the patient, but once detected it might lead to disqualification from driving and thus enormous handicap. Another example is the presence of HIV

infection. Once this pathological finding has been detected, even in the absence of any impairment or disability, it can severely affect the person's life style leading to loss of job, inability to obtain life insurance, and social isolation. The link between disability and handicap is also weak compared with the strong links between handicap and various environmental factors such as social expectations and prejudices, legal framework, family support, physical environment and financial support, which have a major effect upon the final level of handicap (Figure 1.2). For instance, the legal environment determines the degree of handicap arising from the impairment of epilepsy which may be mild and cause little or no impairment or disability and yet prevent driving. Conversely some patients with severe disabilities are able to lead successful lives with minimal handicap.

Within the handicap domain, and indeed within the impairment and the disability domains, abnormality is considered to be a reduction of some pre-existing state that has occurred as a result of the underlying pathology. Unemployment and poor housing are therefore not handicaps by themselves. It is the loss of a job or the deterioration in housing as a result of a disease, which constitutes the handicap.

1.3 Difficulties with the ICIDH model

The ICIDH model which intended to offer a comprehensive framework for the classification of human 'disablement' has succeeded to a great extent in making a very complex problem easily understandable. However practical experience as well as theoretical considerations have shown that the application of the ICIDH in certain situations may be difficult (Bury, 1987; Thuriaux, 1995). Some of these difficulties are semantic. The terms used in this classification were not new and had already been used to describe the consequences of disease in slightly different contexts. The term 'handicapped', for example, is used in the United States to describe people with disability in a pejorative way (Badley, 1993). The three basic constructs of 'disablement' describe closely related theoretical entities and consequently they overlap. The distinction between impairment and disability can be particularly difficult, as in the case of aural, visual, and language impairments and communication disabilities. The distinction between disability and handicap can also be difficult because the description of what constitutes handicap, particularly in

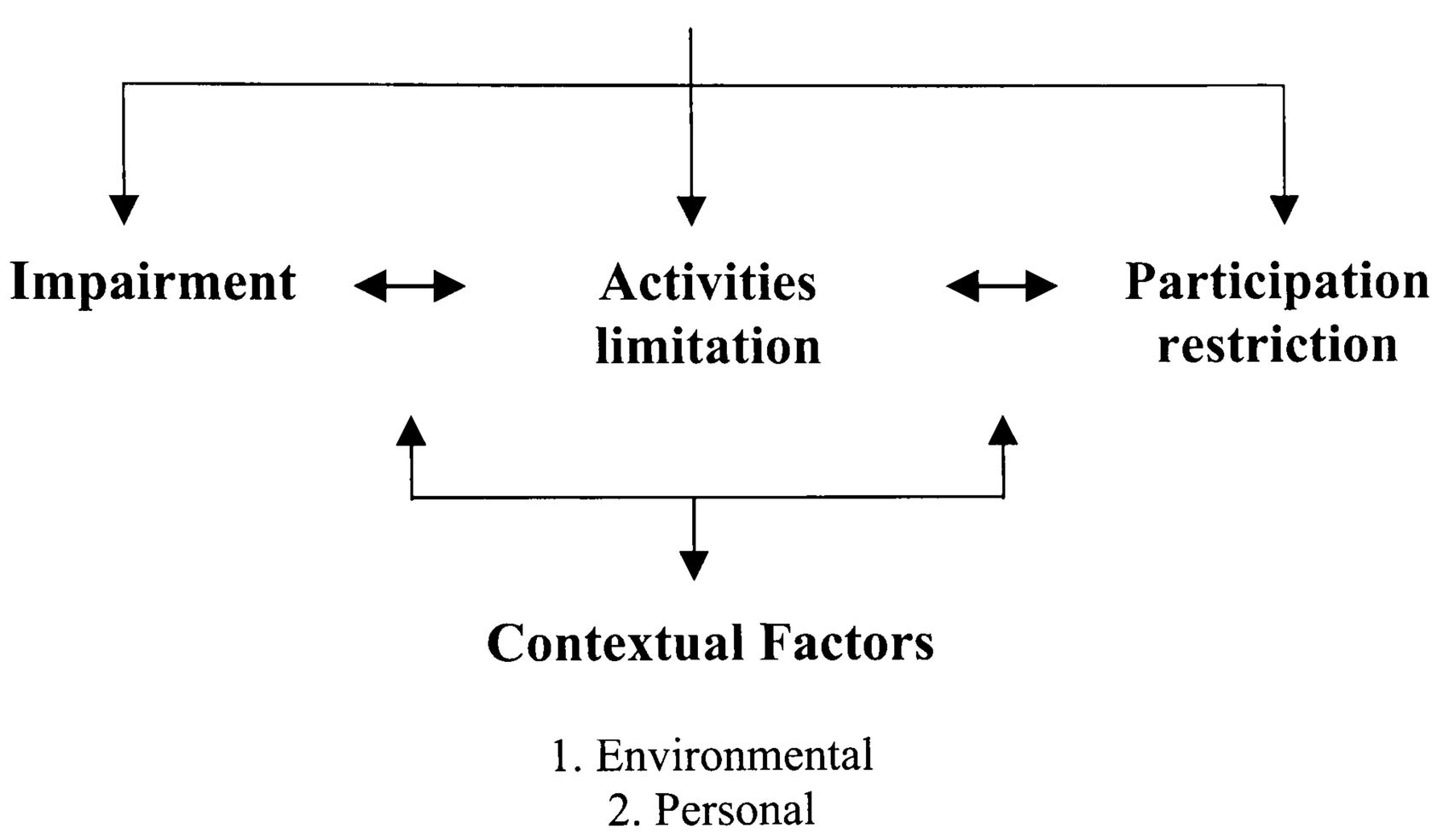
relation to mobility, physical dependence, and orientation, is equally explained in terms of disability. A further source of ambiguity relates to the concept of social role and to the distinction between disability in various activities of daily living, such as housework, laundry, and shopping and the social roles which these activities form. The self-reporting required to assess some of the ICIDH measures creates a potential source of subjectivity. Finally it is always necessary to add the dimension of time to the description of 'disablement' since the ICIDH is a record of a static position.

It is therefore important for this model to be used flexibly, and for its users not to be bound rigidly by it. It should also be remembered that the ICIDH does not classify diseases but attempts to classify situations which are themselves fluid and changing, and excessive concerns with terminology and the problem of overlap should not interfere with the final goal of this model.

In the light of these shortcomings, the WHO has started a global initiative aimed to revise the ICIDH with the help of multiple partners using a consensus-building exercise guided by the present scientific thinking and the practical needs of health science research. These efforts have produced the ICIDH-2, which is currently in a draft format awaiting finalisation in the year 2000 following the end of the ongoing field trials (World Health Organisation, 1997). The revised version has kept the same three-tier structure but employed a slightly modified terminology to avoid some of the problems outlined above. The new terms include impairment, activities limitation (to replace disabilities), and participation restriction (to replace handicap) (Figure 1.3). The ICIDH-2 has also acknowledged the complex interactions between these three basic constructs, and the various environmental and personal contextual factors.

Despite its difficulties, the ICIDH remains the best framework available for health science research. It provides a comprehensive view of long term disease consequences and emphasises the importance of the multidisciplinary approach to patients' care.

Figure 1.3 The ICIDH-2 model illustrating the current understanding of the relation between its three dimensions (World Health Organisation, 1997)



1.4 Quality of life

Quality of life is a universally recognised term whose origin is unknown and whose meaning remains difficult to define. In 1947, the WHO adopted a broad definition of health as:

" a state of complete physical, mental, and social well-being, and not merely the absence of disease and infirmity" (World Health Organisation, 1947).

Two years later, Karnofsky and Burchenal (1949) expanded the criteria for evaluating success in cancer trials to include functional status, mood and general well being. The term 'quality of life' appeared in Index Medicus in 1975, and a specific section was dedicated to it in 1977 (Smith, 1993). Since then, quality of life issues have become steadily more important in health care research and practice.

1.4.1 Definition

Researchers from different disciplines have approached quality of life assessment from different perspectives, and have consequently used different definitions as appropriate for their own work. This is partly due to the fact that the construct of quality of life encompasses not only health-related factors but also many other non-medical aspects of life including standard of living, quality of housing and neighbourhood, job satisfaction, etc. The specific term 'health-related quality of life' has therefore been advocated for use in the context of health science research (Gill, 1995; Testa and Simonson, 1996). An acceptable definition of health-related quality of life is:

"*The functional impact of an illness, and its consequent therapy, upon the patient, as perceived by the patient*" (Schipper et al., 1990).

Health-related quality of life has traditionally been assessed in three principal domains: physical, psychological and social (Smith, 1993). The physical domain is concerned with the effect of disease on patients' abilities to carry out their normal activities of daily living (e.g. daily functioning, pain, vitality and energy). The psychological domain deals with the emotional aspects of disease (e.g. perception of well being, self-esteem, anxiety, and depression). The social domain considers the impact of disease on patients' social activities (e.g. social interactions with family, friends, work colleagues, and within the community) (Devinsky, 1995).

Health-related quality of life measures have an important role in controlled clinical trials in determining the positive impact of treatments, and perhaps more importantly, in assessing the potential negative effect of these treatments on patients' quality of life which other traditional outcome measures fail to capture. They also allow direct comparison between the social and economic burden of different diseases on society and the effect of their treatment. They can inform policy decisions concerning the allocation of health care funds by quantifying what they can buy in terms of quality of life, and thus providing some basis for cost benefit and cost utility comparisons. These measures are also important in obtaining qualitative information on patients' preferences and subjective perceptions, which may have a powerful influence on real life treatment decisions outside clinical trials (LaRocca et al., 1996).

1.4.2 Classification

Health-related quality of life measures can be divided into two types: generic and disease-specific.

A). Generic Instruments

These measures are designed to be broadly applicable across different types and severity of diseases, medical interventions, and demographic and cultural groups, so as to permit comparisons across studies. These measures can be further divided into health profiles and utility measures.

1). Health profiles

These multi-dimensional measures combine several discrete scales, which quantify the major dimensions of health-related quality of life by means of patients' self-reports. They are particularly helpful in providing an assessment of the various areas, which may be adversely influenced by ill health in the form of a composite score. Examples of these measures include the Sickness Impact Profile (Bergner et al., 1981) and its British version the Functional Limitation Profile (Patrick et al., 1985), the Nottingham Health Profile (Hunt et al., 1981), and the Short Form 36 Health Survey Questionnaire (Garratt et al., 1993).

2). Utility measures

These measures are based on the idea that quality of life may be viewed as a uni-dimensional phenomenon. They attempt to gain a single index value of health status by eliciting patients' preferences for their health status. Typically, these measures are presented as a single number, or 'utility', ranging from 0 (death) to 1 (full health), which can be combined with the estimated life duration to derive 'quality-adjusted life years' (QALY's = number of years lived × the quality of life experienced per year). Such measures are potentially useful in understanding relative treatment benefits in both the quantity and quality of life, and in comparing different health care programs by combining the two main effects of therapy, survival and quality of life. Examples of these measures include the Quality of Well Being Scale (QWBS) (Kaplan and Anderson, 1988), the Quality-Adjusted Time without Symptom and Toxicity Scale (Q-TWIST) (Gelber et al., 1992), and the EuroQol visual analogue scale (EuroQol Group, 1990).

B). Disease specific instruments

These instruments concentrate on quality of life issues particular to a specific disease, and are therefore most appropriate for clinical trials in which specific therapeutic interventions are being evaluated. The list of disease specific health-related quality of life instruments is nearly endless (Patrick and Deyo, 1989), but they all share the disadvantage of not allowing for direct comparisons between different diseases. A desirable approach is to supplement a short generic instrument, such as the SF–36, by additional disease specific items to create a comprehensive measure, such as the Multiple Sclerosis Quality of Life - 54 Instrument (Vickrey et al., 1995b), and the Epilepsy Surgery Inventory - 55 (Vickrey et al., 1995a).

1.5 Health-related quality of life and medical ethics

The interpretation of health-related quality of life varies between doctors on one hand, and health economists and administrators on the other. Clinical medicine is concerned with disease sufferers and its ethical perspective is therefore deontological: to do the best for the patient regardless of cost. Public heath concern is to reduce the burden of disease suffered by the population and its ethical standpoint is therefore utilitarian: to do the greatest good for the greatest number (Ebrahim, 1995). Health-related quality of life measures may therefore be used by doctors to decide which of a range of possible treatments will bring about the most favourable 'quality of life'. Ironically the same measures, particularly QALYs, may be used by health care administrators to discriminate between patients in competition for limited resources. The elderly and very sick are the most obvious losers.

Advocators of the utilitarian approach commit a logical error by failing to notice the important distinction between the 'quality' of life and the 'value' of life. They assume that the 'value' of human life is not fixed but varies according to age, ability, and social status. The idea that human beings are unique and have qualities that cannot be measured is not considered. A person may be twenty or ninety, but their existence as a human being is equally important. The assumption that 'if a person's quality of life decreases then the value of this person's life must proportionally decrease' is therefore fundamentally wrong (Seedhouse, 1994). The

uncertainty about the validity and the precision of these measures should also be considered. It is difficult, if not impossible, to be certain about how long people will live, or to judge whether one patient's suffering and pain is the same as another. Health-related quality of life calculation depends on difficult concepts and considerable assumptions, and their interpretation therefore requires cautious consideration.

1.6 Health-related quality of life and the ICIDH

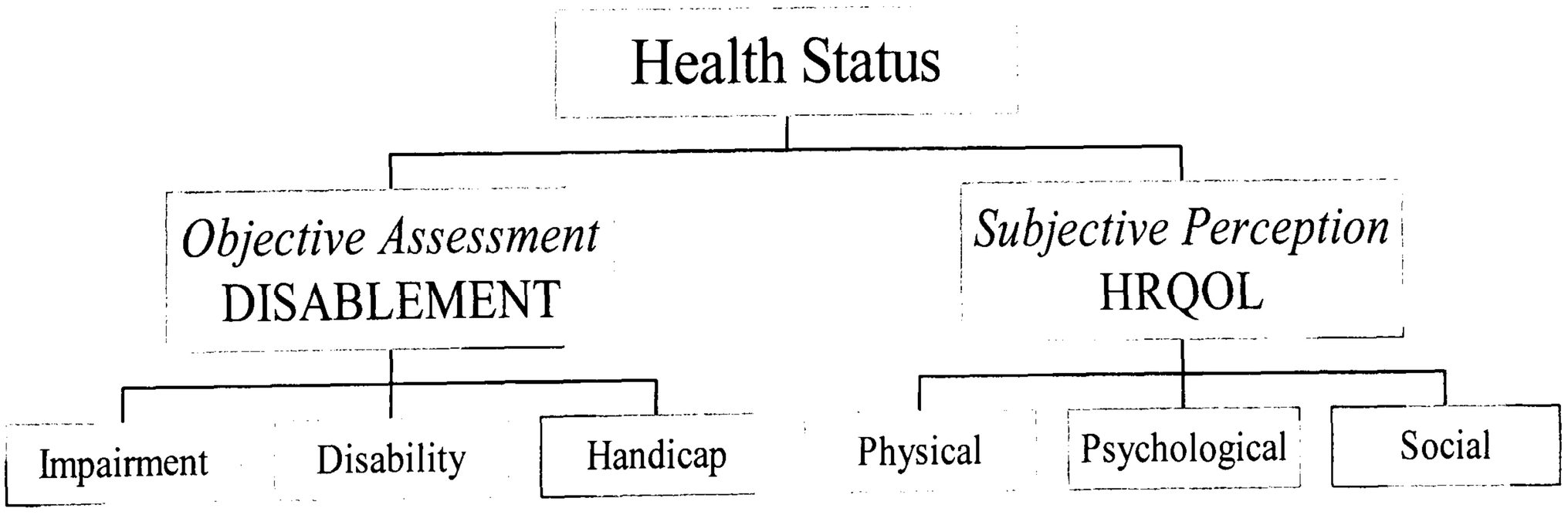
One of the main justifications of using health-related quality of life measures is their intrinsic value in giving a patient-centred, rather than a physician-centred, view of disease consequences. Clinician-oriented outcome measures fail to address the subjectively assessed aspects of health status such as pain, vitality, and general well being, and may therefore provide an incomplete picture of the impact of illness on patients. Theoretically, health-related quality of life measures provide a more uniform method of administration (self-administration) and may therefore reduce observer bias and variability. In reality, health-related quality of life assessment is difficult to standardise because of the multiplicity of ways in which different individuals perceive, respond, and adapt to their illness. These measures are inherently subjective and may therefore be more difficult to be used in patients with psychological or cognitive impairment. This is particularly important since many patients with disabling neurological diseases have considerable deficits in both cognitive skills and emotional control (Wade, 1996). The ICIDH model, on the other hand, has the virtues of simplicity, comprehensiveness and international recognition. It provides an objective assessment of the health status, which is based on directly observable or easily ascertainable dimensions. Its self-reporting elements are not based on value judgements and can therefore be verified against a detailed history obtained from the patients, their families, close friends or carers.

Where does health-related quality of life fit into the ICIDH model? There is no consensus in the literature on how to incorporate these two concepts (Gill, 1995; Wade, 1996; Hobart et al., 1996a). It has been suggested that health-related quality of life might be viewed as the final common pathway of impairment, disability and handicap, or as an 'umbrella' which embodies these three measures (Ebrahim,

1995). In practice, health-related quality of life measures comprise a collection of items which reflect the impairment (pain, anxiety, etc.), disability (inability to walk or climb stairs, etc.), and handicap (inability to keep a paid job, etc.) elements of health as perceived by the patient. In multiple sclerosis, objective impairment and disability assessments by clinicians using the Expanded Disability Status Scale (Kurtzke, 1983) correlate closely with patients' self-assessment of their own physical disablement using the physical functional domain of the Short Form 36 Health Survey Questionnaire (r = 0.87) and the Functional Limitation Profile (r = 0.77), but not with the overall health-related quality of life as measured with EuroQol (Hutchinson and Hutchinson, 1995; Rothwell et al., 1997).

Health-related quality of life and the ICIDH are therefore better viewed as representing the subjective and objective sides of the health status (Hughes and Sharrack, 1998). They offer different but complementary and equally important accounts of health status. By augmenting rather than replacing one another, they could contribute to a more comprehensive outcome assessment (Figure 1.4).

Figure 1. 4 Conceptual scheme representing the relation between the three domains of the ICIDH and health related quality of life



1.7 Conclusion

The comprehensive approach of the ICIDH has provided a successful model for understanding the differences between patients who have the same pathology but different impairments, disabilities, and handicaps. Health-related quality of life measures have incorporated the patients' distinctive values into the assessment process. The ICIDH and health-related quality of life measures represent the objective and the subjective sides of the health status. Together they provide a global assessment of the non-fatal consequences of disease on the affected individuals.

Chapter 2

PATHOLOGY AND DISABLEMENT IN MULTIPLE SCLEROSIS

2.1 Introduction

Multiple sclerosis is a chronic disabling disease which affects approximately 0.1% of Caucasians of north and central European ancestry (Sadovnick et al., 1996). It is the most common cause of chronic neurological disability in young adults (Freeman et al., 1996; Thompson, 1996a). Diagnostic criteria define the age of onset as 10 to 59 years (Poser et al., 1983), but in the majority of cases the onset is between the age of 20 and 40. Females are affected 2 times more often than males (Compston, 1998). Most patients become disabled as the disease progresses, with approximately 50% requiring walking aids or the use of wheelchairs within 15 years of onset (Weinshenker, 1994). Despite this, life expectancy is not substantially altered by this disease. The mean survival is reported to ranges between 25 and 35 years in various series (Poser et al., 1989)

There is presently no cure for multiple sclerosis. Treatment of affected individuals has relied for a long time on supportive care, management of complications such as limb spasticity, bladder instability and infections, together with occasional courses of oral or intravenous corticosteroids during periods of acute deterioration (Ebers, 1994). More recently, a steady flow of new potentially effective agents have emerged with data resulting from phase II and III studies claiming effectiveness in reducing the number of relapses, or modifying the natural history of disease progression (Thompson and Noseworthy, 1996).

2.2 Pathology

The pathological hallmark of multiple sclerosis is the presence of multiple foci of demyelination, or plaques, in the central nervous system. Such plaques have a predilection for certain regions especially the periventricular areas, corpus

callosum, optic nerves, and the spinal cord (Moor, 1998; Lassmann, 1998). Acute plaques are characterised by focal areas of myelin loss with extensive infiltrates of lipid laden macrophages, lymphocytes, monocytes and reactive astrocytes, and intense perivascular cuffs of lymphocytes and monocytes (Raine, 1991). Significant degrees of axonal damage and oligodendroglial proliferation and remyelination are also frequently seen (Trapp et al., 1996; Lassmann, 1998). Chronic active plaques show a gradient of pathological changes. The central regions reflect older events whereas the borders show evidence of ongoing active demyelination and attempts at remyelination (Moor, 1998). Chronic silent plaques are relatively acellular, with complete demyelination, gliosis, and axonal loss (Moor, 1998).

2.3 Pathophysiology

The presence of demyelination, and to some extent inflammation, can cause conduction slowing or block in the affected axons which account for the acute deficits that characterise the early years of the disease (Ffrench-Constant, 1994). Recovery from these acute relapses has several mechanisms including the resolution of oedema and inflammation, remyelination, restoration of conduction in persistently demyelinated fibres by the insertion of new sodium channels into the internodal membrane, and cortical readaptation (Moll et al., 1991; McDonald, 1998). Irreversible deficits may be due to failure of the demyelinated fibres to restore conduction, or axonal loss which accounts for the progressive increase in disability that characterise the later stages of the disease.

2.4 Aetiology

The cause (s) of multiple sclerosis remains unknown. The evidence suggests that environmental, genetic, as well as immunological factors play part in the aetiology. Multiple sclerosis has a non-random geographical distribution with high prevalence rates in temperate latitudes of both northern and southern hemispheres. This gradient distribution often represent genetic variation within the population (Ebers and Sadovnick, 1993), but the seven fold difference between southern New Zealand and Tasmania (high prevalence) and northern Queensland (Low prevalence)

in ethnically similar populations points strongly to environmental factors (Hammond et al., 1988). The nature of these factors remains uncertain.

Evidence for genetic susceptibility comes from three observations. Firstly the prevalence of multiple sclerosis varies between various ethnic groups, being high in Northern European Caucasians and low in Native and Black Americans and Asians even when they live in high prevalence areas (Kurtzke et al., 1979). Secondly the presence of specific Major Histocompatibility Complex class II associations especially with DRw15 and DQw6 in North Europeans (Olerup and Hillert, 1991). Thirdly the high concordance rate in monozygotic twins (25-31%) (Sadovnick et al., 1993; Mumford et al., 1994). The observation that an increased concordance rate is not seen in adoptive siblings also suggests that familial cases are due to shared genetic susceptibility rather than shared environment (Ebers et al., 1995). Several susceptibility genes have been found to play a part, but their nature is not fully elucidated (Compston, 1998).

The evidence that multiple sclerosis is an autoimmune disease is strong but circumstantial. A widely held view is that the immediate cause of the pathological process is an aberrant T-cell mediated immune response to a variety of myelin antigens which circulate in the blood of multiple sclerosis patients (Utz and McFarland, 1994). When activated, helper T-cells cross the blood-brain barrier to the central nervous system, under the influence of adhesion molecules, where they interact with specific antigens presented by the Major Histocompatibility Complex class II molecules on macrophages and astrocytes. This interaction results in cytokine secretion (tumour necrosis factor–α and interferon-γ), T-cell proliferation, B-cell and macrophage / macroglia activation and synthesis of inflammatory mediators causing breakdown of the blood brain barrier (Hartung et al., 1995). Autoantibodies directed to myelin antigens cross the damaged blood brain barrier, or are locally produced by B-cells which have been stimulated by the T-cells, activating the complement system and leading to wide spread oligodendrocyte death and subsequent demyelination (Ffrench-Constant, 1994). Results from therapeutic trials with immunomodulatory agents support this concept since treatments that

augment immune function have exacerbated disease activity whereas immunosupressive drugs have produced modest clinical benefits (Goodkin, 1994).

2.5 Clinical Course

The clinical course of multiple sclerosis is highly variable. It ranges from a fulminating disorder which can be fatal within months, Marburg's disease (Weinshenker, 1994), to an asymptomatic condition which is only recognised incidentally at autopsy (Gilbert and Sadler, 1983) or by magnetic resonance imaging in the asymptomatic monozygotic co-twins of multiple sclerosis patients (Sadovnick et al., 1993). The most commonly observed clinical course is characterised by episodes of acute periods of worsening (relapses, exacerbations, bouts, attacks), gradual progressive deterioration of neurological function, or combinations of both. A recent international survey of 215 leading clinicians involved with multiple sclerosis revealed general consensus about four different clinical courses (Lublin and Reingold, 1996):

A). Relapsing remitting multiple sclerosis

The disease course is characterised by clearly defined relapses with full recovery or with residual deficit upon recovery. Periods between relapses are characterised by the lack of disease progression.

B). Primary progressive multiple sclerosis

The disease course is characterised by a progressive phase from onset with or without occasional plateaus and temporary minor improvements.

C). Secondary progressive multiple sclerosis

The disease course is characterised by initial relapsing remitting phase which is followed by progression with or without occasional relapses, minor remissions or plateaus.

D). Progressive relapsing multiple sclerosis

The disease course is characterised by a progressive phase from onset with clear acute relapses which may be followed by partial or full recovery. Periods between relapses are characterised by continuing progression.

In the majority of cases (60%), multiple sclerosis runs a relapsing and remitting course which culminates ultimately in a secondary progressive phase

(Weinshenker, 1994). In 15-20% of cases, the illness runs a 'benign' course with relatively few attacks early on but without developing any, or with very little, permanent disability. Primary progressive multiple sclerosis accounts for 15% of all cases (Weinshenker, 1994). This variability presents researchers with major problems when assessing whether an apparent improvement in an affected individual represents a true response to a therapeutic intervention, or simply a natural remission which would be experience with the passage of time.

2.6 Diagnostic criteria

A diagnosis of multiple sclerosis is based on evidence of two or more lesions in the central nervous system, dissociated in both space and time. To ensure diagnostic uniformity for clinical research studies, several diagnostic criteria have been suggested. The Poser criteria are currently the most widly used (Poser et al., 1983). According to these criteria, the diagnosis of multiple sclerosis is classified as clinically definite, laboratory supported definite, clinically probable, or laboratory supported probable based on the number of relapses experienced by the patient (one or more), the evidence on clinical examination of abnormal signs suggestive of the presence of one or more anatomically unrelated lesions in the central nervous system, the presence of paraclinical evidence (evoked potentials or magnetic resonance imaging) of central nervous system abnormalities suggestive of demyelination, and the presence of oligoclonal gammaglobulin banding in the cerebrospinal fluid.

2.7 Symptoms and signs

Plaques of demyelination can occur virtually anywhere in the central nervous system and in those cranial nerves (olfactory, optic, and auditory) in which the axons are supported by oligodendrocytes. The spectrum of the clinical features can therefore be extremely diverse although there are some reasonably predictable clinical presentations with variable combinations of visual, motor, sensory, and autonomic abnormalities. The severity of the symptoms and clinical signs is often related to the location rather than to the extent of the pathological process.

2.7.1 Initial symptoms

The onset of multiple sclerosis is mono-symptomatic with symptoms referable to a single anatomical site or system in approximately 50% of patients (Matthews, 1991). Weakness of one or more limbs is present at onset in 20-40% of patients, and this is often accompanied by dysaesthesiae or sensory loss (McAlpine D, 1972; Weinshenker et al., 1989b; Matthews, 1998). The initial symptoms are mainly sensory in about 20-45% of the cases (Weinshenker et al., 1989b; Matthews, 1991). Optic neuritis is present at onset in 17-25% of patients (Shibasaki et al., 1981; Weinshenker et al., 1989b). Limb ataxia and impaired balance occur in about 13% (Weinshenker et al., 1989b), diplopia in about 12% (McAlpine D, 1972; Weinshenker et al., 1989b), and vertigo in about 5% of cases (McAlpine D, 1972). More unusual presenting features e.g. epilepsy, facial palsy, retention of urine, or paroxysmal symptoms are present at the onset is about 5% of the patients and, if in isolation, their significance is often unrecognised (Matthews, 1991).

2.7.2 Symptoms and signs during the course of the illness

The mass of the published material on the clinical features during the course of multiple sclerosis consists largely of lists of the symptoms and signs encountered by patients in different series during the course of the disease. In two of such series, the results were similar.

A). Poser's series

Poser and co-workers reported large series of 1271 hospital patients with a mean age of 31.1 years and a mean disease duration of 11 years (Poser et al., 1979). Signs of upper motor involvement were present in 80% of the patients, weakness in 78%, sensory change in 73%, optic nerve signs in 48%, ocular signs in 14%, brainstem and cerebellar involvement in 77%, cerebral (mainly mental) abnormalities in 36%, and autonomic disturbance, mainly sphincter disturbance, in 56%.

B). Shibaski's series

Shibaski and co-workers published the clinical features of a series of 204 British and 60 Japanese hospital patients, with a mean age of 42 and 39 years and a mean disease duration of 11 and 8 years respectively (Shibasaki et al., 1981). Both

population groups showed the same preponderance of weakness (80% and 80% respectively), paraesthesiae (84% and 77%), sphincter disturbance (74% and 58%), ataxia (48% and 58%), optic atrophy (71% and 70%) and diplopia (39% and 35%). Notable differences were seen in dysphagia (3% and 23%), tonic seizures (4% and 28%) and Lhermitte's sign (15% and 42%).

These and other similar studies (Kurtzke, 1970; Shepherd, 1979) have provided important information on the clinical features of multiple sclerosis, but the great majority were related to hospital-based patients and none has included all potential patients in a defined population. These studies are likely therefore to have missed many patients with mild or severe diseases. An entirely accurate account of the symptoms and signs could only be compiled from series of a large geographically defined population studies of clinically or laboratory definite cases minutely observed over the whole course of the disease. Such studies are not available, and in default of such unattainable data the best alternative is cross sectional surveys of population-based cohorts. Swingler and Compston conducted such a study.

C). Swingler and Compston's series

Swingler and Compston described the frequency and spectrum of morbidity of multiple sclerosis in Southern Glamorgan County in South Wales (population of 376,718) (Swingler and Compston, 1992). A total of 441 patients were identified through thorough case ascertainment, of whom 301 (68%) were seen and assessed. Three hundred and eighteen patients (71%) of the original cohort had clinically or laboratory supported definite multiple sclerosis and 42 patients (9%) had clinically or laboratory supported probable multiple sclerosis. The patients were interviewed and examined using a standard performa which was designed to collect information concerning demographic characteristics, symptoms, signs, the Expanded Disability Status Scale (EDSS) and its Functional Systems (Kurtzke, 1983) and the Ambulation Index (Hauser et al., 1983). The patients had a mean age of 48.7 years (range 10-85), a mean disease duration of 16.5 years, and a mean EDSS score of 5.0.

1). Symptoms

During the month before assessment, weakness was the most common complain having been reported by 80% of the patients, followed by sensory

disturbance, ataxia, bladder symptoms, fatigue, cramps, altered bowel functions, dysarthria, blurred vision, poor memory, diplopia, dysarthria, and vertigo (Table 2.1).

Table 2.1 Frequency (%) of symptoms in the MS population of South Glamorgan ($n = 301$)

Symptom	At any time	At onset	At prevalence	Persistent
Weakness	89	22	80	62
Sensory	87	34	73	52
Ataxia	82	11	72	58
Bladder	71	1	62	45
Fatigue	57	2	48	31
Cramps	52	0.6	44	26
Diplopia	51	8	26	18
Visual	49	13	33	23
Bowel	42	0	37	19
Dysarthria	37	0.6	25	16
Vertigo	36	4.3	19	13
Facial pain	35	2	14	9
Poor memory	32	0.3	27	0
Headaches	30	2	17	7
Neuropsyciatric	23	0.3	16	7
Deafness	17	0.6	13	8
Facial weakness	16	1	5	3
Dysphagia	13	0.3	10	5
Skin sores	12	0	7	4
Blackouts	11	0.6	4	2
Agonise	6	0.3	2	0.3
Others	10	1	8	5

The tendency for impairments to accumulate with time was illustrated by the increasing frequency of complaints between onset and assessment day. Most symptoms had become two and six times more common by prevalence day, but

fatigue, cramp, sphincter disturbance, dysarthria, cognitive, and emotional problems showed disproportionate increases in frequency during the course of the illness. Symptoms of longer duration were recorded most commonly at onset. These symptoms included weakness (14.4 years), diplopia (14 years), sensory disturbance (13.1 years), ataxia (12.5 years), and blurred vision (12 years). Symptoms of shorter duration were commonly associated with the later stages of the disease. These symptoms included memory impairment (6.2 years) dysphagia (6.4 years), pressure sores (6.4 years), bowel dysfunction (6.0 years), bladder disturbance (8.1 years), and spasticity (8.3 years).

2). Signs

The most common finding in this survey was a defect of visual function (92%). Seventy-two percent of the patients had corrected visual acuity in one or both eyes of 6/9 or less, 26% had field defects, and 48% had nystagmus. Twenty-one percent of the patients had dysarthria, 6% had facial weakness and 6% had bulbar weakness and dysphagia. Limb weakness was seen in 74% of the patients ranging from monoparesis (7%) to quadriparesis (26%). Seventy-eight percent of the patients had evidence of sensory abnormality of the limbs.

3). Mobility

Sixty-nine percent of the patients in this survey were able to stand, 48% were able to walk unaided, 30% required a stick, crutches or a walker, 22% were wheelchair bound and only 25% walked with a normal gait. More detailed assessment was obtained by using the Ambulation Index which showed that only 21% were asymptomatic and fully active. A further 28% reported symptoms but were able to walk 25 feet in $\leq$ 20 second, 27% required unilateral or bilateral support and 23% were wheelchair bound half of whom were unable to transfer.

As judged by the EDSS, a total of 22.6% of the patients were either free from the manifestations of the disease or had a minimal deficit, 25% had moderate dysfunction but were able to walk independently, 23% suffered a mixture of moderate and severe impairments and required regular assistance for walking, 16% were essentially wheelchair bound and 13% were severely impaired or restricted to bed.

2.8 Impairment, disability, and handicap

To plan and evaluate the required services for the affected individuals with multiple sclerosis, it is important to characterise carefully the needs of these patients in a population-based cohorts. Few studies have assessed the degree of impairment in such cohorts as detected by the EDSS but without utilising other measures of disability and handicap (Weinshenker et al., 1989a; Weinshenker et al., 1989b; Runmarker and Andersen, 1993). Only two studies have assessed the degree of impairment, disability, and handicap in a comprehensive manner (Rodriguez et al., 1994; Midgard et al., 1996). The results of these two studies were similar, and I will therefore only review the results of Mayo Clinic study (Rodriguez et al., 1994).

In this study, all known cases of multiple sclerosis in Olmsted Country, Minnesota, USA (population of approximately 100,000) were identified via the computerised centralised index at the Mayo Clinic (Rodriguez et al., 1994). A total of 162 patients with a median age of 47.5 (range 17-87), median disease duration of 15.4 years and a median EDSS of 3.5 (range 1-9.5) were identified and assessed. Ninety-four percent of the patients had clinically definite multiple sclerosis. Patients were assessed using the Minimal Record of Disability (International Federation of Multiple Sclerosis Societies, 1985) which included the EDSS as a measure of impairment, the Incapacity Status Scale as a measure of disability, and the Environmental Status Scale as a measure of handicap. The study therefore provided a cross sectional analysis of the level of impairment, disability and handicap in this cohort.

A). Impairment

The results of the neurological assessment using the EDSS Functional Systems are shown in Table 2.2. Thirty-three percent of the patients had marked paraparesis, paraplegia, hemiplegia or quadriparesis (pyramidal scores 3 to 6), 13.0% showed moderate or severe truncal or limb ataxia or severe ataxia of all limbs (cerebellar scores 3 to 5), 12.9% showed severe extra-ocular muscle weakness, dysarthria or dysphagia (brain stem scores 3 to 5), 22.2% showed loss of vibration, proprioception or pain sensation (sensory scores 3 to 6), 24.7% had frequent urinary incontinence, a need for almost constant catheterisation, or needed constant

measures to evacuate stools (bowel and bladder scores 3 to 6), 9.3% had corrected visual acuity worse than 6/36 in either eye (visual scores 4 to 6), 3.7% had severe decrease in mentation or dementia (mental scores 4 and 5) and 27.2% had severe spasticity that resulted in major interference with function (spasticity score 3).

Table 2.2 Neurological impairment (%) in the Mayo Clinic study as assessed by the Kurtzke's Functional Systems (n = 162)

	Grade							
Functional Systems	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**NA**
Pyramidal	15.4	17.9	15.4	17.9	13	16.7	3.7	0
Cerebellar	38.9	16.0	25.9	9.3	3.7	0	-	6.2
Brain stem	45.7	22.8	17.3	8	4.3	0.6	-	1.2
Sensory	30.2	22.8	18.5	15.4	4.9	1.9	0	6.2
Bowel & bladder	27.2	25.9	21.6	8	13.6	0.6	2.5	0.6
Visual	16.7	37	27.2	6.2	3.1	3.1	3.1	3.7
Mental	49.4	29	13	4.3	2.5	1.2	-	0.6
Others (spasticity)	45.7	14.2	13	27.2	-	-	-	-

B). Disability

The results of the constructed interview using the Incapacity Status Scale are shown in Table 2.3. The study showed that 58% of the patients needed assistance with stair climbing or were unable to perform this task (scores 2 to 4). Walking aids, orthoses or wheelchairs were required for locomotion in 41.4 % of the cases (scores 2 to 4). Most patients reported normal or minimal problems with bowel (71%) or bladder function (51.9%) (scores 0 and 1). Human assistance (scores 3 and 4) was reported as necessary for various activities of daily living including bathing (22.8%), dressing (20.3%), grooming (13.6%), and feeding (8.7%). A minority of patients (11.7%) reported being able to read only very large print (score 3) or were legally blind (score 4). Dysarthria interfering with communication (score 2 to 4) was reported by 13.5% of patients. A minority of patients (4.3%) had mood or thought disturbance that required psychotherapy or hospitalisation (score 3 or 4).

Disturbance in mentation enough to interfere with everyday activities (score 2 to 4) were reported by 18.6% of patients. Many patients (43.3%) complained of fatigue troublesome enough to cause impairment of functioning (score 2 to 4). Most patients (62.3%) reported no difficulties with their sexual function.

Table 2.3 Neurological disability (%) in the Mayo Clinic study as assessed by the Incapacity Status Scale (n = 162)

	Grades				
Categories	**0**	**1**	**2**	**3**	**4**
Stair climbing	28.4	13.6	31.5	3.7	22.8
Ambulation	51.9	6.8	13	9.3	19.1
Transfer	50.6	17.9	13.6	6.2	11.7
Bowel function	49.4	22.2	11.1	15.4	1.9
Bladder function	27.2	24.7	21	20.4	6.8
Bathing	43.2	15.4	18.5	11.7	11.1
Dressing	47.5	21.6	10.5	8.6	11.7
Grooming	62.3	19.1	4.9	7.4	6.2
Feeding	56.6	20.4	12.3	5.6	3.1
Vision	45.1	37	6.2	8	3.7
Speech and hearing	55.6	30.9	12.3	0	1.2
Medical problem	54.3	28.4	14.2	1.9	1.2
Mood and thought	39.5	43.8	12.3	4.3	0
Mentation	57.4	24.1	9.9	5.6	3.1
Fatigability	23.5	33.3	27.2	10.5	5.6
Sexual function	62.3	9.3	8.3	8.6	11.1

C). Handicap

Most patients (53.1%) were working full time. Sixty percent of men and 45% of women were employed (97.5% of the patients were employed before disease onset). A minority of the patients (15.4%) identified themselves as unemployed, not having any housework, or not attending school. Most patients (77.1%) were able to maintain their usual financial standard without external support. No or only minor personal assistance was required in 62.4% of the patients. Community services

(including seeing a doctor, nurse, physiotherapist, social worker, and home help) for more than one hour per week were needed by a very small number of patients (19.8%). Most patients reported normal or minimal difficulties with social activities.

This study demonstrated that the functional status of patients with multiple sclerosis was more favourable than previously recognised. Approximately one-third of the patients had marked paraparesis, paraplegia, or quadriplegia, less than one-fourth needed intermittent or almost constant catheterisation for bladder dysfunction, and only few patients had severe cognitive impairment requiring supervision.

2.9 The economic implications of multiple sclerosis

Multiple sclerosis is very costly to the individual, health care system, and society. Studies attempting to calculate the true cost of this disease have considered the direct healthcare costs in terms of hospital care, drugs, and long-term care, and indirect costs in terms of loss of production due to morbidity and premature mortality (Jonsson and Henriksson, 1998). Based on the 1994 data, the annual cost of multiple sclerosis in the United States was estimated at over $34,000 per person, translating into a conservative estimate of national annual cost of $6.8 billion, and a total lifetime cost per case of $2.2 million (Whetten-Goldstein et al., 1998). Much of the cost (57%) was in the form of burdens other than direct personal health care, including loss of earning, the cost of formal and informal care. At least three studies addressing the cost of multiple sclerosis in the UK have been published (O'Brien, 1987; Holmes et al., 1995; Blumhardt and Wood, 1996). Hospital costs were found to dominate the direct cost and in all studies, with the indirect costs far outweighing direct costs (Table 2.6).

The cost of multiple sclerosis, both direct and indirect, increases with the severity of the disease as measured by the EDSS (The Canadian Burden of Illness Study Group, 1998). The burden of this disease in terms of long-term care is also considerable. In a recent study of community care for severely disabled people in the UK, patients with multiple sclerosis were found to receive the highest number of weekly hours of care at home from formal sources such as district nurses, local authority home help, private agencies and voluntary organisations (Phillips, 1995).

Table 2.4 Cost (£ million) of multiple sclerosis in the UK

Cost	**1986 / 1987** [1]	**1993 / 1994** [2]	**1994** [3]
Direct costs	18.2	23.2	73.9
Hospital	14.2	19.0	67.3
GP	1.7	2.1	4.6
Drugs *	2.4	2.1	2.0
Indirect costs	100.0	250.1	395.0
Total (£ million)	118.2	273.2	468.9

Study 1 (O'Brien, 1987) and study 2 (Blumhardt and Wood, 1996) included England and Wales only; study 3 (Holmes et al., 1995) included the entire UK; * Values were taken before the introduction of interferon beta

2.10 Conclusion

Multiple sclerosis is the most common cause of chronic neurological disability in young adults. The disease course is variable and unpredictable although progressive in nature. Classically, it begins with a relapsing and remitting course which progresses subsequently into a progressive phase with a gradual accumulation of wide ranging and often complex disabilities. This results in a major burden of suffering for patients and their families and makes substantial demands on health, social, and voluntary services.

Chapter 3

THE EVALUATION OF NEUROLOGICAL OUTCOME MEASURES USED IN MULTIPLE SCLEROSIS

3.1 Introduction

Outcome measures allow the classification of patients according to the presence and severity of the disease process (pathology), the resulting disablement in terms of the clinical condition of the affected individuals (impairment), their functional capacity (disability / activities limitation), and their social disadvantages (handicap / participation restriction), or according to their subjectively perceived health related quality of life (Hughes and Sharrack, 1998). The choice of specific outcome measures in clinical trials depends on the nature of the study and the research hypothesis being tested, and on the psychometric properties of these measures. In phase II studies, which are designed to assess the biological effects of therapeutic interventions on patients, a measure of pathology or impairment would be appropriate. In phase III studies, which are designed to assess the clinical effect of therapeutic interventions on the functional capacity of patients, a disability, handicap, or health related quality of life measure would be more desirable.

Regardless of their conceptual nature, the practical value of any outcome measure depends on its clinical usefulness and on its scientific integrity (Table 3.1). A clinically useful instrument must be appropriate for the research hypothesis being tested, acceptable to patients and clinicians, practical to administer, and cost effective (Whitaker et al., 1995). As clinical usefulness does not guarantee scientific integrity, such an instrument should also be reliable, valid and responsive (Hobart et al., 1996e).

Table 3.1 Desirable attributes of clinical outcome measures

Desirable attributes	Explanation
Practical attributes	
Ease of administration	The instrument should be easy and quick to administer by any health care personnel
Acceptability	The instrument should be user friendly and should achieve high levels of clinician and patient compliance
Cost effectiveness	The instrument should be economical of time and resources
Performance attributes	
Reliability	The instrument should be internally consistent and able to generate reproducible scores when applied by the same (intra-) or different (inter-rater) observers or by the same patient (test-retest reliability) in the case of self-administered instruments
Validity	The instrument should be able to measure what is intended to be measured
Responsiveness	The instrument should be able to detect clinically significant change over an appropriate period of time
Appropriateness	The instrument should be able to discriminate between patients with differing degrees of disease severity, and should have no 'ceiling' or 'floor' effects

3.2 Reliability

Reliability considers whether an instrument is capable of producing measurements which are consistent, accurate, and reproducible. There are two forms of reliability: internal consistency and score reproducibility.

3.2.1 Internal consistency

Internal consistency assesses the homogeneity of multidimensional scales by measuring the extent to which their items measure the same concept. If items were chosen without regard for homogeneity the resulting scale could end up tapping different traits. However if the correlation between these items is too high, some of them may be redundant. One of the oldest methods of assessing homogeneity is

item-total correlation which assesses the correlation between an individual item and the rest of the scale after omitting this particular item. The usual 'rule of thumb' is that each item should have a correlation coefficient with the total score of > 0.2 (Streiner and Norman, 1995c). Items which do not achieve this target should be discarded. Another approach is the 'split-half reliability' in which the scale is randomly divided into two sub-scales which are then correlated with each other. As there are many ways to divide the items of a scale, Cronbach devised a method, called Cronbach's alpha, which gives an average of all the possible split-half reliabilities of a scale (Cronbach, 1951). The value of Cronbach's alpha should be above 0.70 but not higher than 0.90.

3.2.2 Reproducibility

Reproducibility addresses the stability of the scores when the scale is administered on different occasions by the same (intra-rater reliability) or by two different observers (inter-rater reliability), or by the same patient (test-retest reliability) in the case of self-report instruments. Reproducibility assesses the degree of random error which could be attributed to the measure itself, the person doing the measurement, or to the person being measured. It is necessary therefore to examine each type of these random errors separately for a comprehensive evaluation.

Reproducibility is assessed by examining the correlation between the scores of two independent observations provided that the patient's clinical status has remained stable in between. There has been considerable debate in the literature about the most appropriate statistical method for assessing reproducibility.

A). Pearson correlation coefficient

This coefficient is based on regression analysis (Norman and Streiner, 1993e). It measures the extent to which the relationship between two variables can be described as a straight line, the regression line, regardless of the intercept value. This coefficient does not differentiate between random and bias errors and tends to give high correlation values as long as there is a constant relation between the two variables. It is also affected by the range of the variables so that the wider the range the higher the correlation even if agreement remains the same (Streiner and Norman,

1995b). The Pearson correlation coefficient is therefore inappropriate as a measure of reproducibility.

B). Percentage of agreement

The simplest way to assess agreement is to calculate the proportion of paired responses in which the two observations are identical. Although simple, this method is strongly influenced by the distributions of the observations so that a preponderance of particular values may achieve high agreement by chance alone (Streiner and Norman, 1995b). The method is only appropriate for dichotomous and categorical scales.

C). Bland and Altman method

This method was initially designed to assess the agreement between two outcome measures and is capable of separating rater bias from random error (Bland and Altman, 1986). It illustrates rater bias graphically by plotting the difference between each pair of observations against their mean. It also allows the calculation of the 'limits of agreement' (equal to the mean score difference +/- twice the standard error of measurement, i.e. the 95% confidence intervals of the mean score difference), and the 'repeatability coefficient' (1.96 times the standard deviation of the mean score difference). The latter coefficient, which has been used by the British Standard Institute as a measure of the reliability of scientific measurements, is an indication of the maximum score difference required to achieve 95% rater agreement (British Standards Institution, 1979).

D). Kappa coefficient

This coefficient was suggested by Cohen to correct for the effect of chance agreement on rater reliability by examining the proportion of total agreement responses in relation to those agreement responses which will be expected by chance only (Cohen, 1960). This coefficient is only appropriate for dichotomous and categorical scales, or ordinal scales which have a small number of possible scores.

E). Weighted kappa / interclass correlation coefficient

Kappa coefficient only considers total agreement and does not give any credit to observations which differ by only one or two points and is therefore inappropriate for continuous scales which have a large number of possible scores. Cohen suggested a statistical method which focuses on partial agreement by using a

weighting system which accounts for the amount of discrepancy between each pair of observations (quadratic weights) giving a partial agreement corrected reliability (Cohen, 1968). This coefficient is practically equal to intraclass correlation coefficient which is calculated from the analysis of variance (Shrout and Fleiss, 1979).

Kappa, weighted kappa and intraclass correlation coefficients scores are interpreted conventionally as: < 0 poor agreement, 0-0.2 slight agreement, 0.21-0.4 fair agreement, 0.41-0.6 moderate agreement, 0.61-0.8 substantial agreement, 0.81-1 almost perfect agreement (Landis and Koch, 1977). As reliability estimates are population dependent, 95% confidence intervals can be constructed for both kappa (1.96 times the standard error of kappa) (Norman and Streiner, 1993b), and intraclass correlation coefficients (Fleiss and Schrout, 1978).

3.3 Validity

To determine that an instrument is measuring what is intended to be measured, some evidence of validity is required. Validity therefore considers the relation between the concept being measured and the instrument used to assess this concept. There are three main approaches to assessing validity.

3.3.1 Face and content validity

This form of validity addresses the extent to which an instrument is representative of the conceptual domain it is intended to cover. Evidence for this type of validity is commonly obtained by comprehensive review of the literature, consensus expert opinion, quantitative patient interviews, and by examining existing measures of the same or different concepts. The evidence for face and content validity is mainly logical, but statistical analysis can also be undertaken to assess the extent to which groups of experts or patients agree with or disapprove of a particular instrument.

3.3.2 Criterion related validity

Evidence for this type of validity is provided by examining the correlation between the instrument and a gold standard. The difficulty with this approach is

usually related to the lack of gold standards which is what justifies the development of the new scale in the first instance. There are two approaches for assessing criterion related validity which refer to whether the instrument is being compared with a gold standard at the same time (concurrent criterion validity) or in the future (predictive criterion validity).

3.3.3 Construct validity

In the absence of a gold standard, validity is established through a series of strategies to examine the relation between the instrument and other measures or behaviours. In practice, evidence for construct validity is gathered by undertaking a series of studies to determine:

A). Convergent validity

The extent to which the measure correlates with measures of related entities.

B). Discriminant validity

The extent to which the measure does not correlate with measures of different entities.

C). Group difference and Hypothesis testing

The extent to which the measure is able to detect differences in groups of patients known to differ in the concept being measured, or to support various hypotheses generated from theoretically based conceptions.

The process of determining construct validity therefore depends of the accumulation of all three types of evidence.

Validity is assessed using Pearson's and Spearman rank correlation coefficients for interval and ordinal scales respectively. Correlation coefficients of 0.35-0.49 have traditionally been interpreted as weak, those of 0.50–0.79 as moderate, and those of 0.80 or greater as strong (Sharrack et al., 1999c).

3.4 Responsiveness

Responsiveness addresses the ability of an instrument to detect clinically significant change over a relatively short period of time. As the ultimate goal of most therapeutic interventions is to induce change in patients' health status, responsiveness is an essential requirement of all outcome measures. Responsiveness

can be determined in several ways including serial administration of the instrument at different times when clinical change is expected to occur such as before and after treatment of known efficacy or through comparison with other criteria of change such as staff and patient perceptions of change. Responsiveness is assessed using Student-t test (Norman and Streiner, 1993a) or Wilcoxon Signed Rank test (Norman and Streiner, 1993f) for continuous and ordinal data respectively. It is also assessed using effect size calculated by dividing the difference between the scores of the first and the second assessments by the standard deviation of the first assessment scores (Kazis et al., 1989). Effect size values have traditionally been interpreted as: <0.19 unresponsive, 0.2-0.49 small, 0.5-0.79 moderate, and 0.8-1 large (Cohen, 1977).

3.5 Appropriateness

Appropriateness refers to the ability of an instrument to discriminate between patients with differing degrees of disease severity (Streiner and Norman, 1995a; van der Putten et al., 1999). It is assessed by examining the instrument's mean / median, score range / standard deviation, and its 'floor' and 'ceiling' effects. The mean / median indicates the central tendency of the scores and should ideally lie near the midpoint of score range. Score range / standard deviation indicates the extent to which the instrument is able to demonstrate variability between subjects. When applied to large populations, the scores of a desirable scale should have a near normal distribution with no 'floor' or 'ceiling' effects. 'Floor' and 'ceiling' effects are calculated as the percentage of the patients scoring the minimum and the maximum possible scores respectively. Values exceeding 20% are considered to be significant (Holmes and Shea, 1997).

3.6 Clinical disablement scales used for multiple sclerosis

The need for a scoring system to assess the effect of experimental interventions in multiple sclerosis has prompted many researchers to develop different clinical rating scales. Many of these scales were developed well before the WHO published the ICIDH (World Health Organisation, 1980), although some of the more recent scales have incorporated this terminology in their framework. None of these scales has been universally accepted, and only a few have been widely adopted for clinical trials.

3.6.1 The Expanded Disability Status Scale (EDSS)

The EDSS has become the best known and the most widely used scoring method in multiple sclerosis (Kurtzke, 1983). It combines impairment and disability in a 20-step ordinal scale which ranges between 0 (normal status) and 10 (death due to MS). In this scale, patients are graded according to the history and the findings of a standard neurological examination in the appropriate grades of a complementary set of eight Functional System scales which include: pyramidal, cerebellar, brain stem, sensory, bowel and bladder, visual, cerebral and 'others'. An overall EDSS score is obtained by combining the different Functional System scores with the patient's ability to ambulate, use their upper limbs, communicate, and swallow. The lower EDSS grades (0 to 3.5) depend largely on the Functional Systems, but the higher grades are determined by the degree of ambulation (4.0 to 7.5), upper limb dysfunction (8.0 to 8.5), or bulbar dysfunction (9.0 to 9.5). The EDSS can only be administered by trained staff, usually neurologists.

A). Face / content validity and appropriateness

Despite its popularity in clinical trials, the EDSS has many problems. The term EDSS itself is a misnomer since the scale rates a mixture of impairment (in the lower grades) and disability (in the higher grades). Subjective variables, particularly in relation to ambulation, play a major role in allocating the final scores. The differences between grades 5.5, 5.0, 4.5 and 4.0 depend on the ability to walk 100, 200, 300 or 500 metres respectively without aid or rest, which are often estimated and not measured objectively. It is often difficult for most patients and many neurologists to estimate these distances accurately and for both parties to reach a mutual agreement on them (Sharrack and Hughes, 1997). The scale lacks precision in defining some of its Functional System grades due to the use of vague terms such as 'mild', 'moderate', or 'severe' which are open to different interpretations. Combining the results of the Functional Systems to allocate patients to the appropriate EDSS grades can also be difficult for patients with high scores on the Functional Systems but relatively normal ambulation. Significant relapses do not necessarily affect grading as ambulation-dependent EDSS scores are not affected by relapses which do not alter the ambulation status of patients. The scale is rather

insensitive to cognitive and upper limb dysfunction, or to the severity of items incorporated in the two-grade Functional System for 'others' such as fatigability, vertigo, pain, or oscillopsia. Cross sectional studies have shown the EDSS to have a bimodal distribution with paucity of patients in the middle value grades (Willoughby and Paty, 1988; Goodkin et al., 1989; Rodriguez et al., 1994). Goodkin and co-workers (1989) studied 425 patients with multiple sclerosis and found that 45.9% scored between 1.0 and 3.5, 21.2% scored between 6.0 and 6.5, and only 4.7% scored between 4.0 and 5.5. Progression in the EDSS is non-linear as patients progress faster between steps 1 to 5 than between steps 5 to 7 (Myers et al., 1992). Ellison and co-workers (1994) have proposed a strategy to cope with this non-linearity in clinical trials by defining worsening as a change of 1.0 EDSS units (i.e. two 0.5 steps) maintained for 90 days for patients with an entry score of 1.0 to 5.0, but 0.5 EDSS units (i.e. one 0.5 step) if the entry score is 5.5 to 7.0 units.

B). Reliability

At least eight reliability studies of the EDSS have been conducted (Noseworthy, 1994). These studies reported high inter- and intra-rater reliability of the Functional Systems scores with 97–100% rater agreement when allowing a difference of 2 points (Amato et al., 1988; Noseworthy et al., 1990; Goodkin et al., 1992), fair to substantial inter-rater reliability of the EDSS (kappa coefficient 0.32-0.76) (Amato et al., 1988; Noteworthy et al., 1990; Francis et al., 1991), and moderate to almost perfect intra-rater reliability of its lower (1.0 to 3.5) grades (frequency of perfect agreement 50-60%, intraclass correlation coefficient 0.88-0.96) (Goodkin et al., 1992).

C). Responsiveness

Ellison and co-workers (1993) found the Disability Status Scale (the previous version of the EDSS) to be insensitive to worsening of patient's clinical status as judged by the assessing neurologist. Hobart and co-workers (1996d) have also found the EDSS to be unresponsive in a group of 64 patients with moderate to severe disability (EDSS 5.0-9.0).

D). Construct validity

The face validity of the EDSS as a measure of combined impairment and disability is confirmed by its high correlation with the Scripps Neurological Rating

Scale (r = -0.84 to -0.89) (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995), and patients' self-assessment of disability using the physical functioning domain of the SF-36 (r = -0.87) (Rothwell et al., 1997), and the Barthel Index (r = 0.89) (Hobart et al., 1996c). The EDSS has also been found to strongly correlate with physical disability as measured by non-medically qualified assistants using the disability questionnaire of the Office of Population Censuses and Surveys (OPCS) (r = 0.84) (Martin et al., 1988; Rothwell et al., 1997). The OPCS disability instrument was developed as generic measure of disability which can be administered by non-medically qualified personnel for the use the 1985 survey of disability among adults.

Despite its shortcomings, the EDSS has been used as an outcome measure in almost all published clinical trials of multiple sclerosis. It has also been used extensively in studying the natural history of multiple sclerosis (Weinshenker, 1994). The use of this scale in future clinical trials is therefore mandatory until such a time that a suitable alternative is devised and universally accepted.

3.6.2 The Scripps Neurological Rating Scale (SNRS)

The SNRS is a 22-item ordinal impairment scale which converts the standard neurological examination into a numerical score using a 3-level scoring system (Sipe et al., 1984). A normal individual receives the full score of 100 points with progressive loss of points for mild, moderate or severe impairment until the worst possible score of 0 point. The SNRS can only be administered by trained staff, usually neurologists.

A). Face / content validity and appropriateness

The SNRS does not adequately reflect cognitive dysfunction. The total points allocated for each neurological system are weighted arbitrarily with high scores for the visual, sensory, motor and cerebellar systems, and low scores for mentation and mood, tendon reflexes and plantar responses. The scale lacks precision as guidelines for defining the degree of impairment (mild, moderate, or severe) of its various items are not provided. This stricture applies particularly to items such as visual fields, optic discs, pupils, and eye movements in which impairment has not traditionally been graded. The SNRS has been reported by Koziol and co-workers (1996) from the

Scripps clinic to have a near normal frequency distribution. This observation has never been re-evaluated by other independent groups.

B). Reliability

Koziol and co-workers (1996) reported high inter-rater agreement with a weighted kappa coefficient of 0.83 and 0.85% score agreement when allowing a difference of 10 points, and high intra-rater reliability with weighted kappa coefficients for the two examiners of 0.98 and 0.99. No other reliability studies by independent groups have been published.

C). Responsiveness

Koziol and co-workers (1996) reported that score changes of the SNRS are 'more gradual' in comparison with those on the EDSS, suggesting that the SNRS is not sensitive to clinical change. There are no other published reports in the literature addressing the responsiveness of the SNRS.

D). Construct validity

The face validity of the SNRS as an impairment measure is supported by its high correlation with the EDSS (r = -0.84 to -0.89) (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995).

3.6.3 The Ambulation Index (AI)

The AI is a semi-quantitative scale which converts ambulation-related disability into an ordinal scale based on the speed and assistance needed for walking 25 feet by specifying 10 grades between 0 (normal status) and 9 (wheelchair-bound and unable to transfer independently) (Hauser et al., 1983). The scale is simple and can be administered by any health care personnel.

A). Face / content validity and appropriateness

Compared with the EDSS, the AI provides a more precise measure of ambulation. However it usefulness as an outcome measure in clinical trials is limited by its mono-dimensional nature and its inability to take into account the wide range of disabilities experienced by patients with multiple sclerosis. The AI has a bimodal distribution with a paucity of scores 7 and 8 (Goodkin et al., 1989; Swingler and Compston, 1992)

B). Reliability

The AI has moderate to substantial inter-rater reliability (kappa coefficient 0.5-0.72) with 95% score agreement when allowing for a difference of 1 point (Francis et al., 1991). No intra-rater reliability studies have been reported in the literature.

C). Responsiveness

Although the AI detected more clinical change than the EDSS in at least one clinical trial (British and Dutch Multiple Sclerosis Azathioprine Trial Group, 1988), there are no published reports in the literature addressing its responsiveness.

D). Construct validity

The face validity of the AI as a disability scale is supported by its high correlation with the EDSS (Herndon and Goodkin, 1997).

3.6.4 The Functional Independence Measure (FIM)

The FIM is an 18-item ordinal disability scale which rates the level of assistance required to perform various activities of daily living using a 7 level scoring system with sum scores ranging between 124 (normal status) and 18 (totally dependent) (Hamilton et al., 1987; Keith et al., 1987a). The FIM can be administered by any health care personnel. However it is a somewhat cumbersome scale which requires reference to a 48-page instruction book and training for its administration.

A). Face / content validity and appropriateness

The FIM does not account for the whole range of disabilities experienced by patients with multiple sclerosis by failing to cover visual, speech, swallowing, affective, or sexual disabilities. The scale's 7 point scoring system lacks precision in some of its items. The ambulation item gives a score of 6 for 'modified mechanically assisted independence', which means that the same score will be given to patients who need unilateral or bilateral support. The same score will also be given to someone who walks slowly but independently, whether or not gait is abnormal. In a study of 201 patients with moderate to severe disability (EDSS 5.0 to 9.0), van der Putten and co-workers (1999) found the FIM sum score to have small 'ceiling' and 'floor' effects.

B). Reliability

Given the multidimensional nature of this scale, the internal consistency of its items is surprisingly very high (Cronbach's alpha 0.94-0.95) suggesting that some items may be redundant (Brosseau and Wolfson, 1994; Streiner and Norman, 1995c). Previous reliability studies showed high inter-rater reliability of the FIM sum scores with intraclass correlation coefficients of 0.83-0.96; high inter-rater reliability of the motor (intraclass correlation coefficients of 0.95-0.97) and the cognitive domains sum scores (intraclass correlation coefficients 0.84-0.88); and variable item score inter-rater reliability (intraclass correlation coefficients 0.14-0.98) (Brosseau and Wolfson, 1994; Hobart et al., 1996b). There are no published reports addressing the intra-rater reliability of this scale.

C). Responsiveness

The FIM sum score is responsive to clinical change thereby enhancing the usefulness of this scale in clinical trials of MS (Hobart et al., 1996d).

D). Construct validity

The face validity of the FIM as a disability measure is supported by the high correlation between this scale and the EDSS and the burden of care (Granger et al., 1990a; Hobart et al., 1996c).

3.6.5 Other scales

A). The Minimal Record of Disability

In 1985, The International Federation of Multiple Sclerosis Societies committee on rating systems suggested a Minimal Record of Disability for multiple sclerosis which was based on the ICIDH and comprised of three scales: the EDSS as an impairment scale, the Incapacity Status Scale as a disability scale and the Environmental Status Scale as a handicap scale (International Federation of Multiple Sclerosis Societies, 1985).

The Incapacity Status Scale is a 16-item ordinal scale with a 5-point scoring system from 0 (normal function) to 4 (loss of function). The scale is administered by patient interview and describes patients' performance in stair climbing, ambulation, toilet/chair/bed transfer, bowel function, bladder function, bathing/dressing, grooming, feeding, vision, speech and hearing, medical problems, mood and thought disturbances,

mentation, fatigability and sexual function. Like other activity of daily living scales, this scale is biased towards upper and lower limb dysfunction which are reflected in more than one of its items.

The Environmental Status Scale is a 7-items ordinal scale with a 6-point scoring system from 0 (normal) to 5 (worst dysfunction). The scale rates the social and environmental impact of multiple sclerosis on patients in terms of actual work status, financial/economic status, personal residence/home, personal assistance required, transportation, community services and social activity.

Both the Incapacity Status Scale and Environmental Status Scale can be administered by allied health professionals or trained volunteers. To date only the EDSS has been adopted and widely used in clinical trials.

B). The Cambridge Multiple Sclerosis Basic Score (CAMBS)

The CAMBS is an ordinal scale which rates the individual contributions of disability, relapse, disease progression, and handicap by using four separate sub-scales and a 5-level scoring system giving a four component score for each assessment (Mumford and Compston, 1993). The CAMBS has been found to be reproducible, with high correlation between its disability component and the EDSS and between its handicap component and both the Barthel Index and the Nottingham Health Profile (Mumford and Compston, 1993). This scale was not designed as an outcome measure for clinical trials or as a substitute for existing scales, but as a useful shorthand record for clinical neurological practice and retrospective case note analysis.

C). Composite scores

In response to the need to develop a new scoring system for multiple sclerosis, the U.S. National Multiple Sclerosis Society convened a task force to develop recommendations for optimal assessment measures for use in future multiple sclerosis clinical trials (Rudick et al., 1996a). The task force proposed a new measurement approach based on the use of quantitative functional composites which consist of simple quantitative functional measures combined into a single score. In the absence of such an ideal composite, the task force recommended the use of a three-tier composite (timed 25-foot walk, the nine-hole peg test, and paced auditory serial addition test) as a secondary outcome measure in future clinical trials whilst awaiting the development of more refined composites (Rudick et al., 1996b).

Unfortunately the proposed multidimensional composite consists primarily of impairment measures which cover only one aspect of the disablement process and are therefore incapable of providing a comprehensive appraisal of patients' health status. These measures also have clear 'ceiling' effects and are not applicable to patients with advanced multiple sclerosis who are unable to walk, manipulate fine objects with their hands or process mental arithmetic (Sharrack and Hughes, 1999b).

3.7 Conclusion

Many neurological rating scales have been proposed to assess the impact of multiple sclerosis on patients, but none has been universally accepted. The EDSS has been the most widely used despite its problems. It combines impairment and disability and is heavily weighted toward ambulation. The SNRS attempts to quantify impairment as measured by the traditional neurological examination. However this and other impairment scales lack direct relevance to patients' functional health status. The AI is a simple and reproducible scale, but it only measures limited aspects of the wide range of disabilities encountered in multiple sclerosis. Current scales of disability and activities of daily living, such as the FIM are not comprehensive to the type of dysfunction which occurs in multiple sclerosis. The recently suggested composite outcome measures comprise mainly impairment scales which have clear ceiling effects. The need for a new outcome measure is recognised.

Chapter 4

THE RELIABILITY OF DISTANCE ESTIMATION

4.1 Introduction

The assessment of patients' walking ability is a simple and practical method of evaluating the functional health status in neurological, respiratory, cardiovascular and peripheral vascular diseases. Such assessments correlate well with more sophisticated measurements of cardiorespiratory function or muscle strength (Sinclair and Ingram, 1980; Bernstein et al., 1994), and are important in assigning scores in many clinical rating scales including the EDSS and the FIM.

The two most commonly used methods are the assessment of the maximum distance that a patient can walk, or the distance that they can walk till the onset of symptoms. These distances are hardly ever measured in everyday clinical practice. Doctors have traditionally relied on their own or their patients' estimates of the distances walked around familiar places. A previous study assessing the accuracy of trained and untrained observers in estimating target distances ranging between 600 to 1550 metres showed wide variability (Fine and Kobrick, 1983). There were no published studies assessing the accuracy of distance estimates made by doctors and patients (Sharrack and Hughes, 1997).

4.2 Subjects and methods

I sent a standard questionnaire to all the consultants at Guy's Hospital explaining the aim of the study and asking them to estimate (in yards or metres) the dimensions of a hospital ward, and the distances between 5 familiar sites in and around the hospital. A category for 'don't know' was provided to prevent guessing. One hundred and five questionnaires were returned (return rate of 53%), of which 100 were completed. The same questionnaire was given to 100 consecutive adult

patients from a general medical / neurology hospital ward and a neurology outpatient clinic in the same hospital. No help or clarification was provided to any of the patients, and none of them had any overt psychiatric disorder or cognitive dysfunction. All study sites were later measured with an architect's tape measure. Data were tabulated and analysed using SPSS 7.5 for windows.

4.3 Results

The consultants were more familiar with the hospital sites than the patients. The number of consultants giving estimates for the six distances varied between 45 and 97 and the number of patients between 10 and 62 (Table 4.1 and 4.2). Both consultants and patients were inaccurate at estimating distances. Their mean estimates correlated moderately with the measured distances (r = 0.73 and 0.56 respectively), and the range of their estimates was very wide and generally greater for consultants. The estimates for the whole group differed by up to 14.6 fold from the measured distances, and the difference between the minimum and the maximum estimates was up to 62.5 fold. This wide variability was partly due to the presence of a few outliers (Figure 4.1) since the differences between the measured distances and the median estimates of both groups were relatively small.

When estimates were expressed as percentages of the measured distances, the estimates of the shorter distances were more inaccurate than those of the longer distances. The patients' mean estimate of a ward 7.2 yards wide was 19 yards, an error of 163.9 %, while the consultants' mean estimate of the same ward was 11 yards, an error of 52.8%. On the other hand, the patients mean estimates of a 349 yard walk to the local station was 495 yards, an error of 41%, while the consultants' mean estimate of the same distance was 371 yards, an error of only 6.3%. However the differences between the mean estimates of both groups were not statistically significant, with the exception of the 349 yard distance (p = 0.023).

Table 4.1 Estimates of distances (in yards) by consultants (n = 100)

Measured distance	Number of estimates	Median (range) of estimates	Mean (SD) of the differences *
7.2	45	9 (4 to 75)	4.3 (11.1)
20.6	46	30 (15 to 300)	24.6 (46.2)
45.6	95	40 (8 to 500)	7.3 (55.6)
126.9	94	100 (11 to 450)	- 14.7 (67.2)
140	97	120 (24 to 547)	- 1.4 (90.5)
349	94	300 (100 to 3282)	21.8 (375.1)

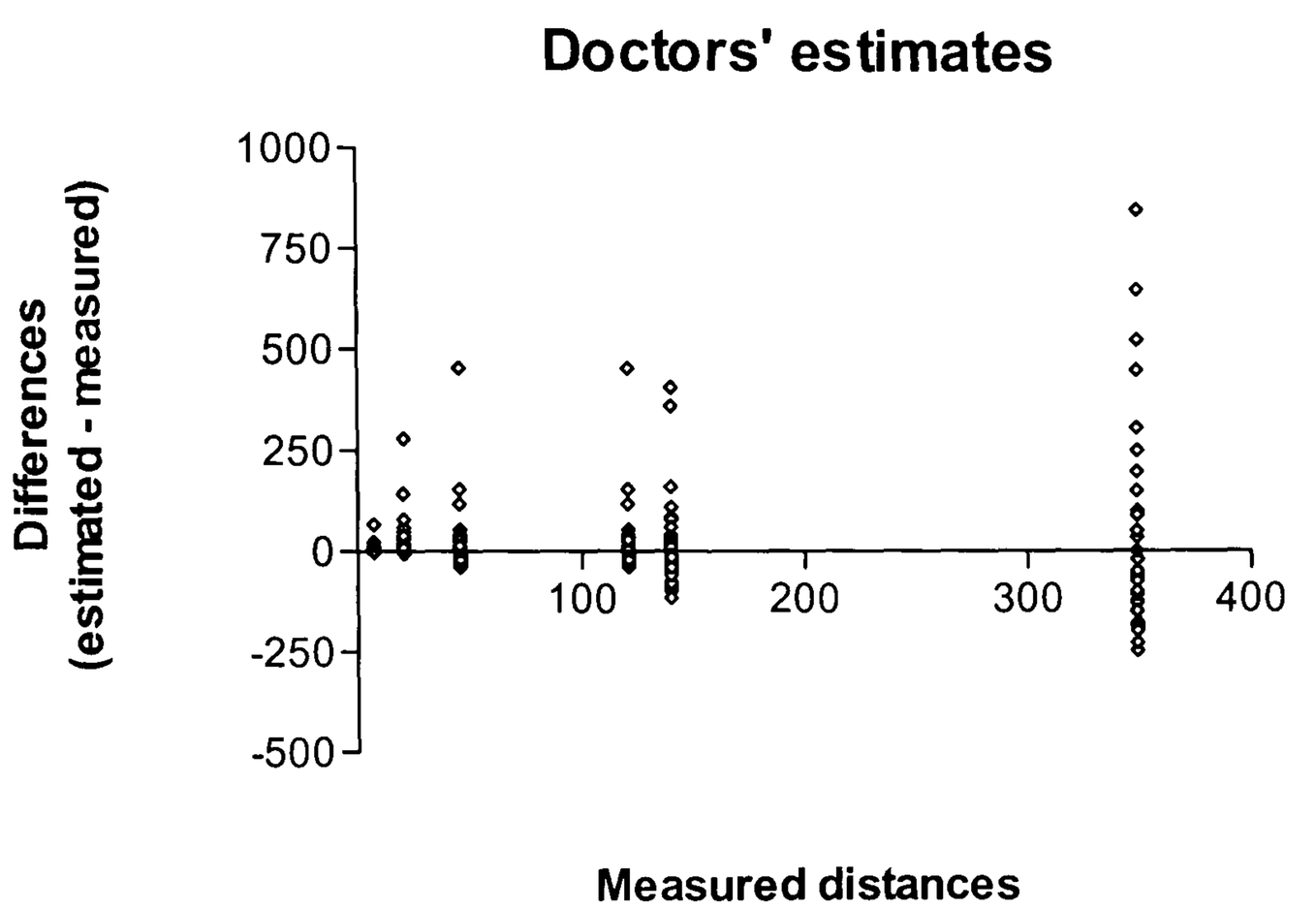
* Between estimated and measured distances.

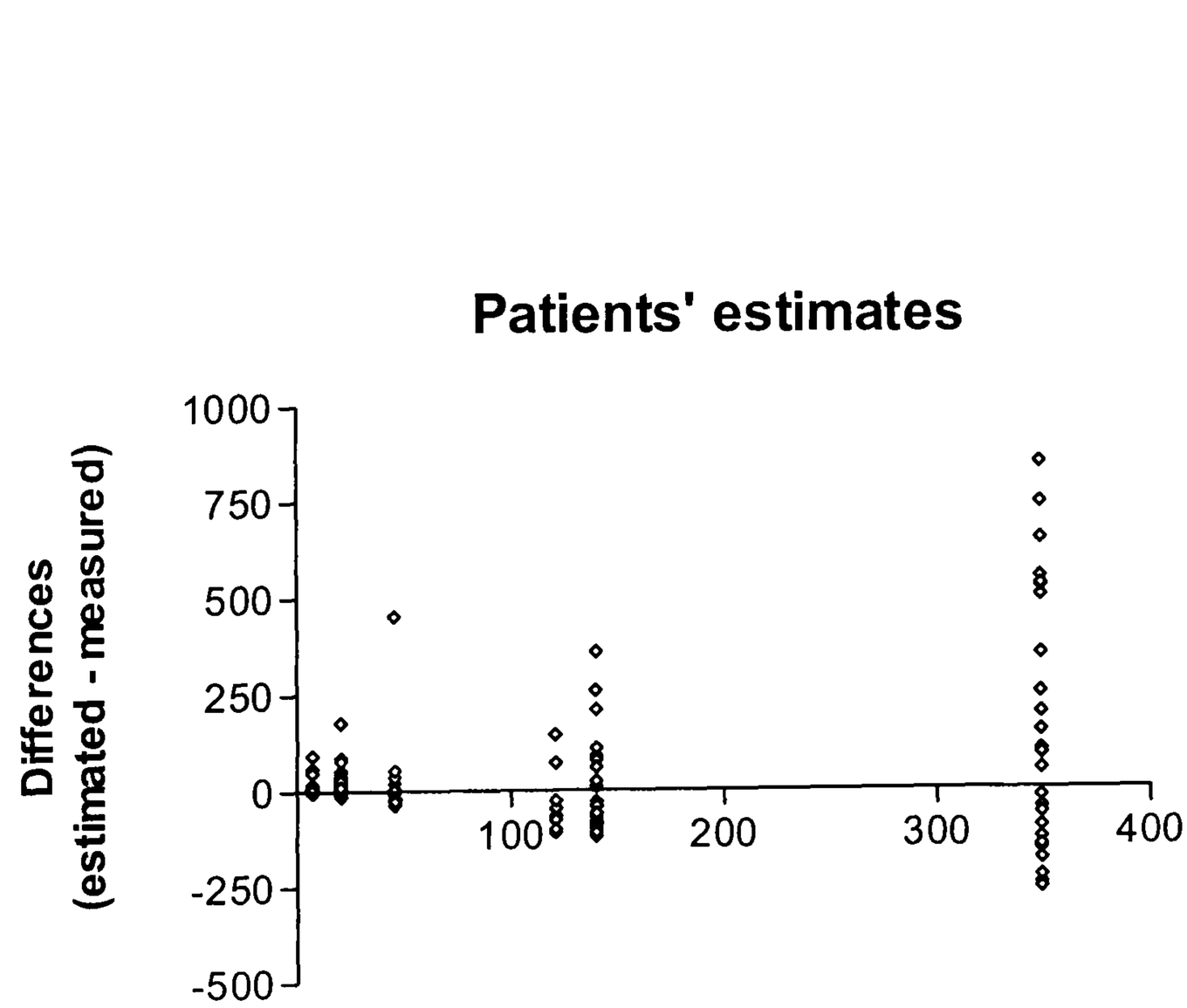
Table 4.2 Estimates of distances (in yards) by patients (n = 100)

Measured distance	Number of estimates	Median (range) of estimates	Mean (SD) of the differences *
7.2	31	10 (4 to 100)	13.5 (26.5)
20.6	30	33 (8 to 200)	25.5 (39.7)
45.6	15	27 (10 to 500)	24.9 (121.6)
126.9	10	70 (18 to 247)	- 33.7 (83.2)
140	62	164 (21 to 500)	26.6 (109.6)
349	58	440 (88 to 1200)	146.3 (287.8)

* Between estimated and measured distances.

Figure 4.1 Differences between estimated and measured distances (in yards) for consultants (n = 100) and patients (n = 100)





4.4 Discussion

This study suggests that people are inaccurate at estimating distances and that medical education is no safeguard. The range of estimates was very wide suggesting that decisions about health status based on individual distance estimates are unreliable. Although the estimates were proportional to the measured distances, indicating that consultants and patients were capable of comparing distances and therefore possibly able to estimate changes, the potential value of this observation needs to be evaluated against the intra-rater variability of these estimates which was not addressed in this study. The comparable inaccuracy of both groups suggests that selecting of patients with mainly neurological disorders (90%) did not bias the results or limit their generalisability to other patient groups. Participants are also unlikely to have deliberately provided spurious estimates since the outlier values were from participants who had given more reasonable estimates for other distances.

The implications of these results are unmistakable. Clinical assessments and therapeutic decisions are often based on such estimates. For example the severity of angina, claudication, and chronic respiratory failure, and the effect of treatment on these conditions is usually assessed by the distance a patient can walk before the onset of symptoms. The results of some of the most hailed clinical trials in multiple sclerosis have been based on ambulation biased clinical rating scales including the EDSS (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995; Jacobs et al., 1996). A 1.0 step change on this 20 grade scale is regarded as a significant change, however the difference between grades 5.5, 5.0, 4.5, 4.0 is the ability to walk 100, 200, 300, or 500 metres respectively. Such distances are usually estimated and are rarely measured objectively in hospital wards or outpatient clinics.

The economic implication of distance estimation is considerable. There are over 1.2 million claimants in the UK in receipt of the higher rate mobility component of the Disability Living Allowance currently £33.9 per week (Department of Social Security, personal communication). This allowance is paid to five categories of patients including 'people who have difficulty with walking' (Steadman, 1993). The eligibility of a person to this allowance is based solely on a

self-assessment questionnaire without the need for any objective examination (Steadman, 1992).

4.5 Conclusion

Assessing the maximum distance that a patient can walk is a simple way of evaluating their functional health status. These distances are usually estimated and rarely measured. I conducted a study to assess the accuracy of distance estimation by asking 100 patients and 100 doctors to estimate 6 distances around Guy's Hospital. Both doctors and patients were inaccurate at estimating distances. The ranges of their estimates were wide suggesting that health related decisions based on individual distance estimates are unreliable.

Chapter 5

THE PSYCHOMETRIC PROPERTIES OF CLINICAL RATING SCALES USED FOR MULTIPLE SCLEROSIS

5.1 Introduction

Clinical rating scales allow the physician to classify patients according to their degree of impairment, disability / activities limitation, handicap / participation restriction, or quality of life, assist in predicting the course of the illness, and provide tools to monitor the response to experimental treatments. Over the last forty years, more than 15 different clinical rating scales have been devised and used in MS research (Sharrack and Hughes, 1996). The scales most commonly used are Kurtzke's EDSS and its related Functional Systems (FS) (Kurtzke, 1983), the Scripps Neurological Rating Scale (SNRS) (Sipe et al., 1984), and the Ambulation Index (AI) (Hauser et al., 1983). The generic Functional Independence Measure (FIM) (Hamilton et al., 1987; Keith et al., 1987a) and the Cambridge Multiple Sclerosis Basic Score (CAMBS) (Mumford and Compston, 1993) have also been proposed as potentially useful clinical scales (Noseworthy, 1994). It is surprising that despite the wide use of these scales in clinical research, data related to their psychometric properties in terms of reliability, responsiveness, validity, and appropriateness remain incomplete. Such data are of paramount importance for assessing the results of previous clinical trials and for designing future trials. This study was designed to assess the reliability, responsiveness, construct validity and appropriateness of these five commonly used scales in MS research. Face and content validity of these scales have already been reviewed in chapter 3.

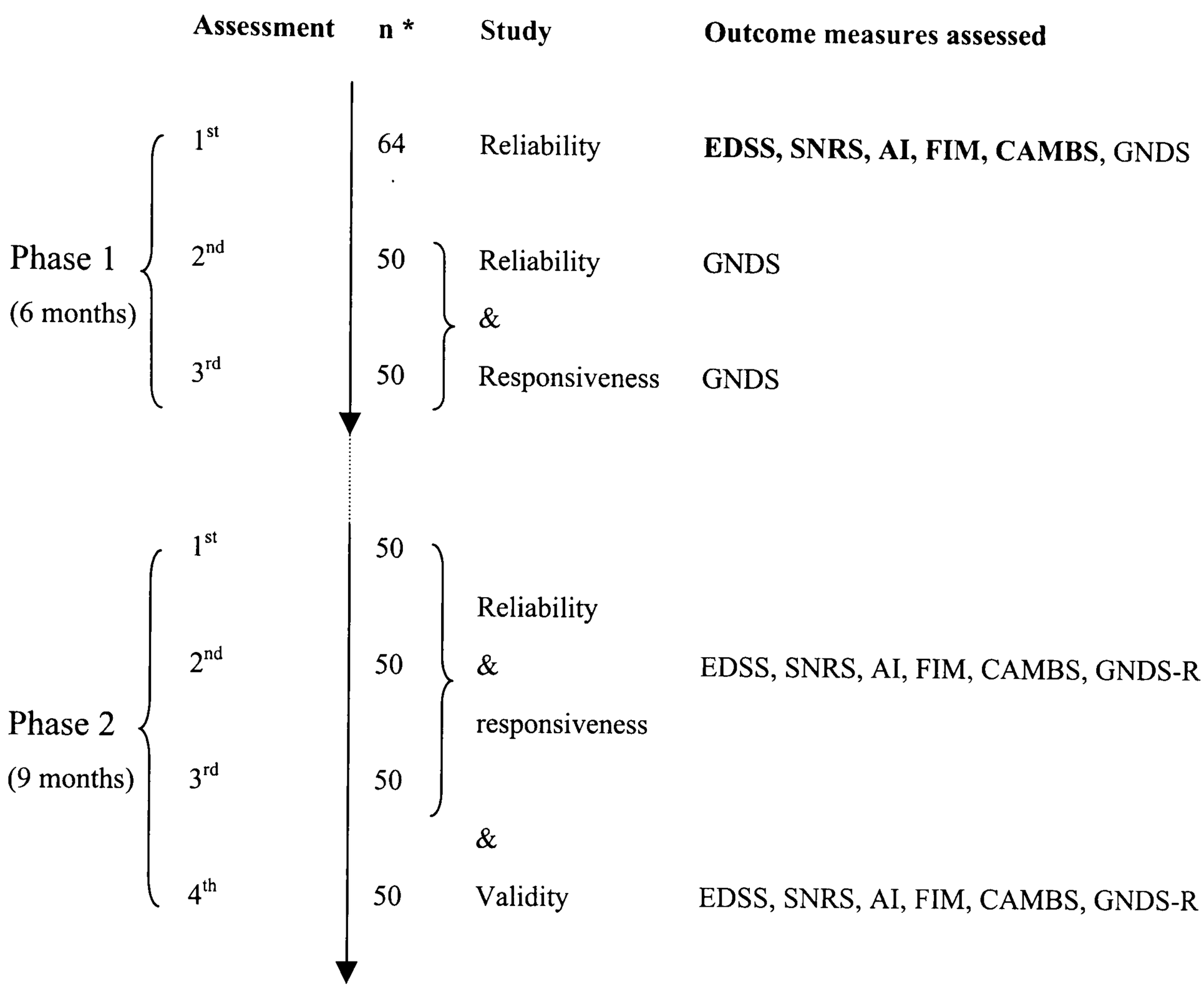
5.2 Patients and Methods

This study was performed in the MS research clinic at Guy's Hospital, London, and was approved by the ethics committee of the local health authority. All subjects consented to take part in the study, and were only recruited if they had clinically or laboratory supported definite relapsing remitting or secondary progressive MS. Sixty-four adult patients were recruited. These patients consisted of 25 patients taking part in a multi-centre randomised double-blind placebo-controlled study of interferon beta 1a in relapsing remitting multiple sclerosis (PRISMS Study Group, 1998), 25 patients taking part in a multi-centre randomised double-blind placebo-controlled study of interferon beta 1a in secondary progressive multiple sclerosis (Paty, 1999), and 14 patients in long-term residential care.

The 50 trial patients were recruited from the general neurology outpatient clinics at Guy's Hospital and other teaching and general district hospitals in and around London. Patients with relapsing remitting multiple sclerosis were eligible if they were aged 18-55 years, had a baseline EDSS score of 0-5.0 and a recorded history of a least two relapses in the preceding two years. Patients with secondary progressive multiple sclerosis were eligible if they were aged 18-55 years, had a baseline EDSS score of 3.0-6.5 and a recorded history of a least two relapses or 1.0 point (or more) increase in EDSS in the preceding two years. The exclusion criteria included the use of immunosuppressive or immunomodulatory treatment during the three months before entry into the study, pregnancy or breast feeding in female patients, and the presence of serious intercurrent illnesses such as cancer, uncontrolled epilepsy, or decompensated liver disease. The 14 patients with more advanced disabilities (EDSS 7.0-9.5) represented all the patients with multiple sclerosis in a local nursing home and at the time of recruitment.

This study overlapped with the two ongoing interferon beta 1a studies (in which patients were assessed every three months) and with the Guy's Neurological Disability Scale study which will be discussed in chapter 6 (Figure 5.1).

Figure 5.1 Overall study design in relation to the ongoing interferon beta 1a trials and the GNDS / GNDS-R studies



* *n* = number of patients in the study

5.2.1 Inter-rater reliability study

Sixty-four adult patients were recruited for this study. These patients consisted of a cohort of 50 patients attending a multiple sclerosis research clinic and 14 patients in long-term residential care. Patients were assessed by three raters, two neurologists (including myself) and a neurology research nurse, who were familiar with the clinical scoring scales used in multiple sclerosis from experience in previous clinical trials and teaching workshops. To standardise the methods of applying the various scales by the three raters, training sessions were conducted prior to the beginning of this study, during which 10 subjects were examined and

scored jointly. Each patient in this study was assessed in the same session independently by the three raters (Figure 5.1). All patients were allocated scores on the EDSS, FS, SNRS, and the AI by the two assessing neurologists, and scores on the FIM and the CAMBS (50 patients only) by one neurologist (myself) and the neurology research nurse.

5.2.2 Intra-rater reliability and responsiveness study

The three raters followed a cohort of 50 MS patients attending the Guy's Hospital MS research clinic for nine months with assessments every three months. During each visit, patients were asked to compare their clinical condition with how they felt on the previous occasion, and indicate whether their condition had since worsened, remained stable, or improved. At the same time, one of the neurologists (myself), who had assessed the patients on the previous occasions, subjectively designated their clinical status as worse, stable, or better. Patients' overall status were later classified as stable, improved, or worsened, if both the patients' and the neurologist's assessments were identical indicating no change, improvement, or worsening respectively. Patients' overall status were otherwise designated as 'uncertain', and all related data were excluded from the final analysis. Patients were also asked to complete the EuroQol health related quality of life questionnaire (EuroQol Group, 1990), and were assigned scores on the EDSS, SNRS, and AI by one neurologist and scores on the FIM, CAMBS, and the Barthel Index (Mahoney and Barthel, 1965) by the other neurologist (myself). In the absence of a gold standard for assessing clinical 'stability' and 'change', and in accord with the methodology of previous studies, intra-rater reliability was tested on the pairs of assessments between which patient's overall status were judged to have remained stable, whereas responsiveness was tested on the pairs of assessment between which they had changed (improved or worsened) (Deyo et al., 1991; Ellison et al., 1993).

5.2.3 Validity study

The validity of the five scales was assessed in the same cohort of 50 patients who took part in the intra-rater and responsiveness study described above. During their third visit, all patients were asked to complete the London Handicap Scale

(Harwood et al., 1994) and the Short Form 36 health survey questionnaire (SF-36) (Garratt et al., 1993), and were ranked by myself according to their ability to work, do their housework, and look after themselves. They were also ranked independently by two raters (myself and the research nurse) according to their subjectively perceived degree of disability. Convergent and discriminant construct validity were tested by assessing the degree to which each scale in this study correlated with the other four scales, and with other measures of disability (Barthel Index), handicap (London Handicap Scale), and health related quality of life (SF-36). Group differences construct validity was assessed by testing the extent to which the scores of these scales correlated with the severity of disability as judged by the two raters. Hypothesis testing construct validity was assessed by testing the hypothesis that scores on any impairment, disability, or handicap scale should be more abnormal in patients who were unable to work or do their housework because of multiple sclerosis, and in patients who were dependent on others for some or all of their activities of daily living.

5.2.4 Blinding

The majority (78%) of the patients in the inter-rater reliability study and all the patients in the intra-rater reliability, responsiveness, and validity studies were taking part in a double-blind therapeutic trial in which the two neurologists were the 'examining' and the 'treating' physicians, and the research nurse was the 'trial co-ordinator'. In this trial, the 'examining' neurologist was responsible for assessing the patients relapse status and assigning scores on the various clinical scales, the 'treating' neurologist was responsible for the overall medical management of the patients, and the research nurse was responsible for the administrative aspects of the study. To comply with the required blinding for both the ongoing therapeutic trial and the current study, the raters refrained from discussing the patients' clinical conditions amongst themselves, and none of them had access to their own or the other raters' previous scores which were kept separate from the patients' clinical records. To reduce the effect of patients' bias on the inter-rater reliability which may result from practice effect or fatigue, no fixed order for the examination of the patients by each rater was observed. Data for the intra-rater reliability study were

collected at three monthly intervals to reduce raters' and patients' bias, which may result from recall of previous assessments.

5.2.5 Statistical analysis

Data were tabulated and analysed using SPSS 7.5 for Windows. Two tailed tests were used for all statistical analyses. The EDSS, SNRS, FIM, AI, CAMBS, Barthel Index, and the disability ranks were treated as ordinal scales, whereas the London Handicap Scale and the SF-36 were treated as interval scales. Reciprocal, logarithmic and square root transformations of the ordinal and skewed data were performed and found to be unhelpful. Descriptive statistics were used to describe the study population in terms of demographic, disease characteristics and score distributions. Inter- and intra-rater reliabilities were assessed with kappa (Cohen, 1960) and intraclass correlation coefficients (Shrout and Fleiss, 1979). Ninety-five percent confidence intervals were constructed for both kappa and intraclass correlation coefficients. The reliability was also expressed as the mean and 95% confidence intervals of inter- and intra-rater score differences to estimate rater bias and the repeatability coefficient as discussed in chapter 3 (Bland and Altman, 1986). Internal consistency of the two multidimensional scales (SNRS and FIM) was assessed using Cronbach's alpha (Cronbach, 1951). The individual contribution of the various scale items to the sum score of the two multidimensional scales was assessed using factor analysis (Norman and Streiner, 1993d). Responsiveness was assessed using Wilcoxon Signed Ranks test and effect size (Kazis et al., 1989). Construct validity was assessed using Pearson's and Spearman rank correlation coefficients for interval and ordinal scales respectively.

5.3 Results

5.3.1 Inter-rater reliability and appropriateness

Sixty-four patients with a wide spectrum of disabilities, ranging from being asymptomatic to being bedridden and completely dependent, were recruited for this study. The group consisted of 42 women and 22 men with a median age 40 years (range 22-74), and median disease duration of 13 years (range 2-35). Inter-rater reliability of the CAMBS was assessed on a subgroup of 50 patients (31 women and

19 men) with a median age of 36 years (range 24–51), median EDSS score of 4.5 (range 0-7.5), and a median disease duration of 12 years (range 2-17).

A). EDSS

The median (range) scores of the two raters were identical at 5.5 (0-9.5). The frequency distribution of the two score sets was bimodal with fewer patients scoring at EDSS 4.0 and 7.0 (Figure 5.2a). Inter-rater agreement on the different Functional System scores was variable with kappa coefficients ranging between 0.41 and 0.67 (moderate to substantial), intraclass correlation coefficients ranging between 0.81 and 0.95 (almost perfect), and repeatability coefficients ranging between 1.2 and 1.6 points (Table 5.1). The largest score differences between the two raters were 2 points for the pyramidal, cerebellar, bladder and bowel, and mental Functional Systems, and 3 points for the brain stem, sensory, and visual Functional Systems. Inter-rater agreement on the EDSS scores was 69%, 89%, 96%, and 100% when agreement was defined as no difference, a difference ≤ 0.5 point (one 0.5 EDSS step), ≤ 1.0 point (two 0.5 EDSS steps), and ≤ 1.5 points (three 0.5 EDSS steps) respectively (Figure 5.3a), with a repeatability coefficient of 0.9 points, a kappa coefficient of 0.65 (substantial) and an intraclass correlation coefficient of 0.99 (almost perfect) (Table 5.1).

B). SNRS

The median (range) scores of the two raters were similar at 69.5 (0-100) and 67 (0-100). The frequency distribution of the two score sets was positively skewed to the 'normal' end of the scale with a smaller cluster at the 'severely impaired' end of the scale (Figure 5.2b). Inter-rater agreement on the different scale items was variable with kappa coefficients ranging between 0.30 and 0.72 (fair to substantial), intraclass correlation coefficients ranging between 0.54 and 0.93 (moderate to almost perfect), and repeatability coefficients ranging between 1.1 and 4.6 points (Table 5.2). Inter-rater agreement on the sum scores was 14%, 59%, 85%, 97%, and 100% when agreement was defined as no difference, a difference ≤ 5 points, ≤ 10 points, ≤ 15 points, and ≤ 19 points respectively (Figure 5.2b), with a repeatability coefficient of 12.1 points, and an intraclass correlation coefficient of 0.97 (almost perfect) (Table 5.3b).

C). FIM

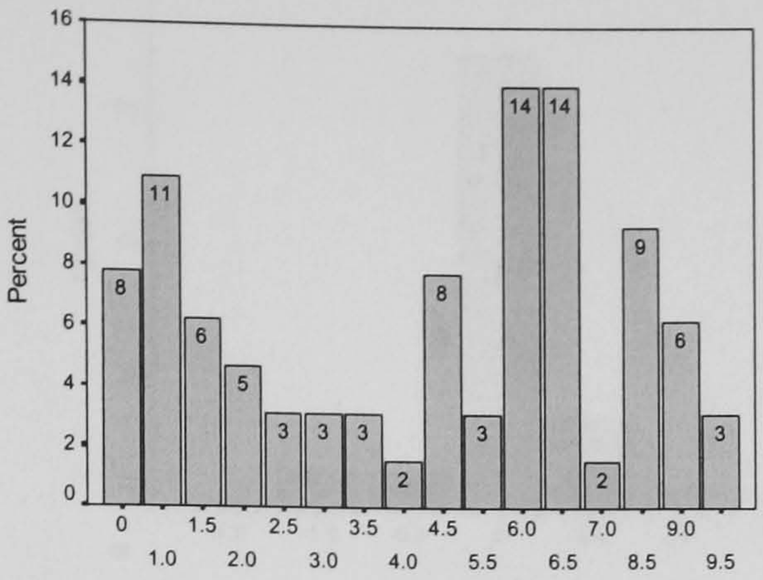
The median (range) scores of the two raters were almost identical at 119 (18-126) and 119 (27-126). The frequency distribution of the two score sets was positively skewed to the 'less disabled' end of the scale with a smaller cluster at the 'severely disabled' end of the scale (Figure 5.2c). Inter-rater agreement on the different scale items was variable with kappa coefficients ranging between 0.26 and 0.88 (fair to almost perfect), intraclass correlation coefficients ranging between 0.56 and 0.99 (substantial to almost perfect), and repeatability coefficients ranging between 0.5 and 2.7 points (Table 5.3). Inter-rater agreement on the sum scores was 25%, 86%, 95.2%, and 100% when agreement was defined as no difference, a difference ≤ 5 points, ≤ 9 points, and ≤ 13 points respectively (Figure 5.3c), with a repeatability coefficient of 8.1 points, and an intraclass correlation coefficient of 0.99 (almost perfect) (Table 5.3).

D). AI

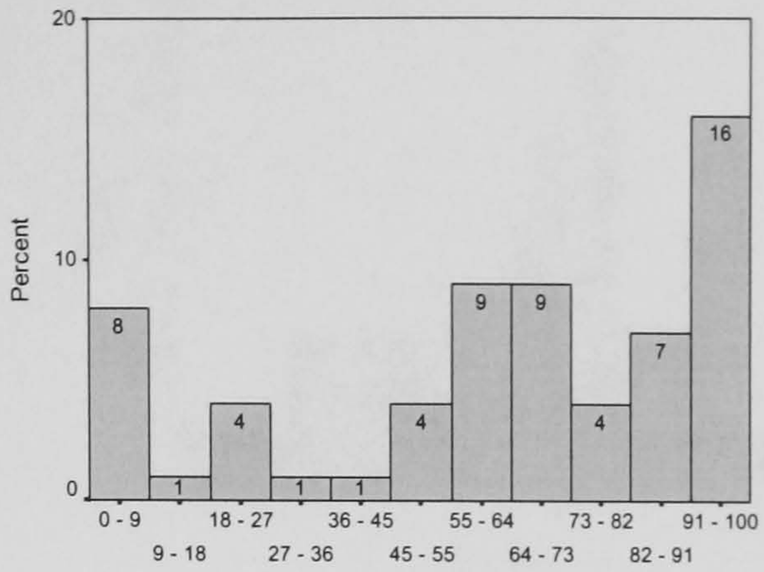
The median (range) scores of the two raters were similar at 2 (0-9) and 3 (0-9). The frequency distribution of the two score sets was bimodal with more patients scoring at the 'normal' end of the scale and fewer patients scoring between 7 and 8 (Figure 5.2d). Inter-rater agreement was 77%, and 100% when agreement was defined as no difference, or a difference ≤ 1 point (Figure 5.3d), with a repeatability coefficient of 1 point, a kappa coefficient of 0.73 (substantial), and an intraclass correlation coefficient was 0.96 (almost perfect) (Table 5.4).

E). CAMBS

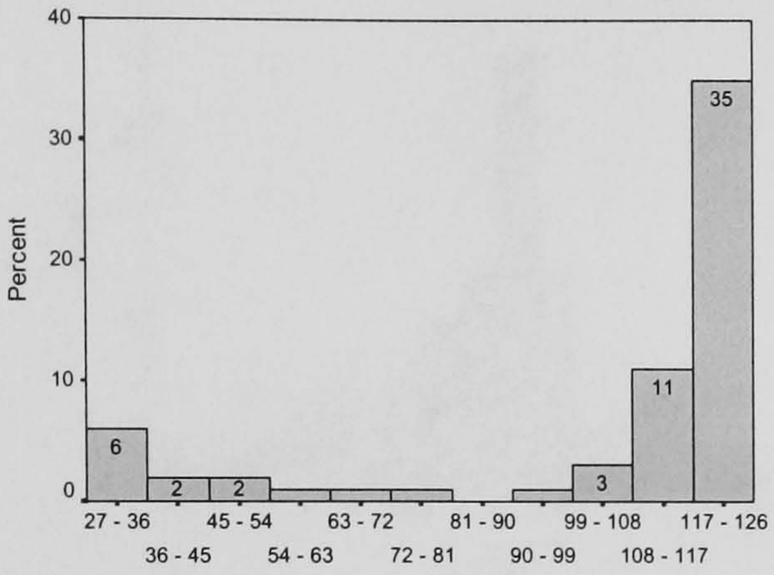
The median (range) scores of the two assessments for the scale's four domains were similar: disability 2 (1-4) and 3 (1-4); relapse 1 (1-3) and 1 (1-4); progression 1 (1-3); and handicap 2 (1-4). The frequency distribution of the two relapse and progression domain score sets was skewed to the 'normal' end of the scales (Figure 5.2f and 5.2g). In comparison, the frequency distribution of the disability domain score sets was skewed to the 'severely disabled' end of the scale, whereas the handicap domain scores were evenly distributed (Figure 5.2e and 5.2h).



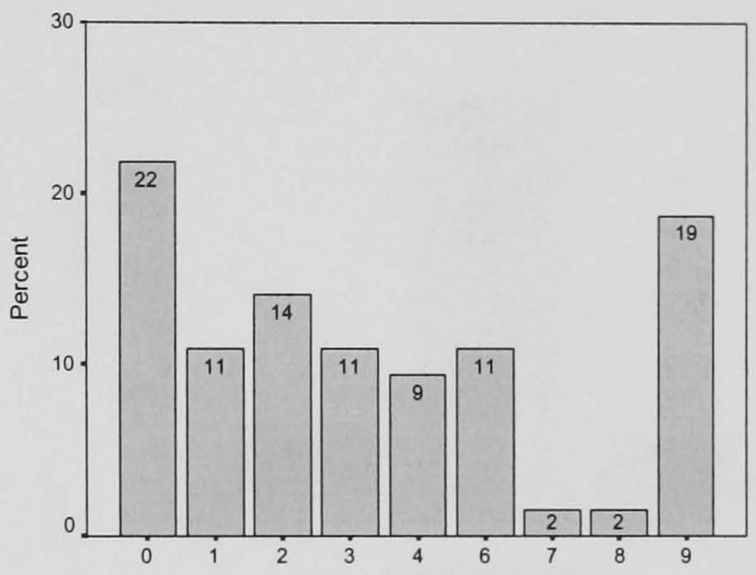
a. EDSS



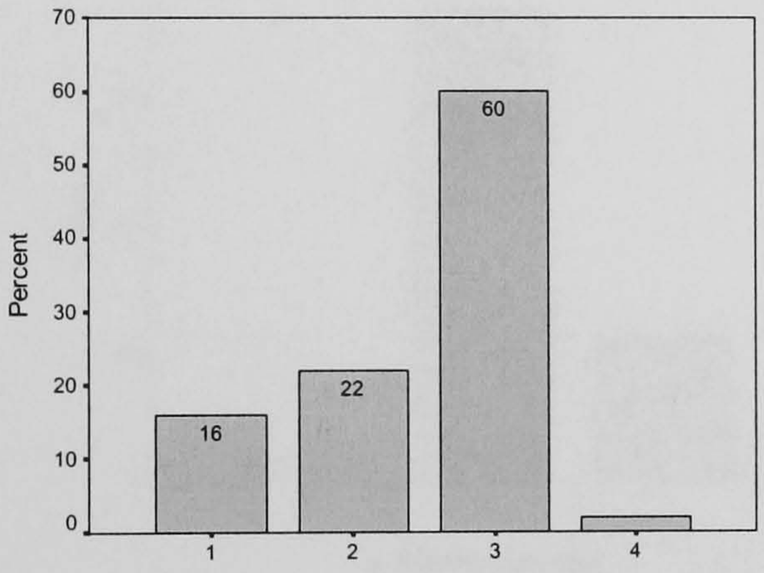
b. SNRS



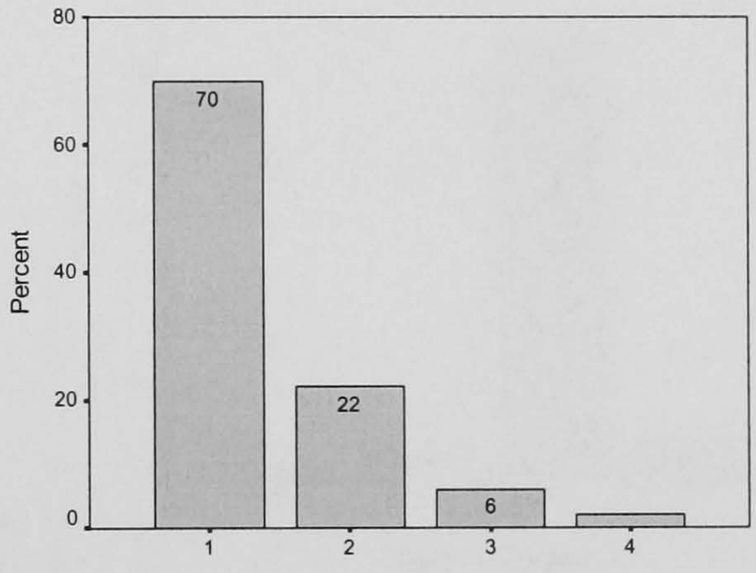
c. FIM



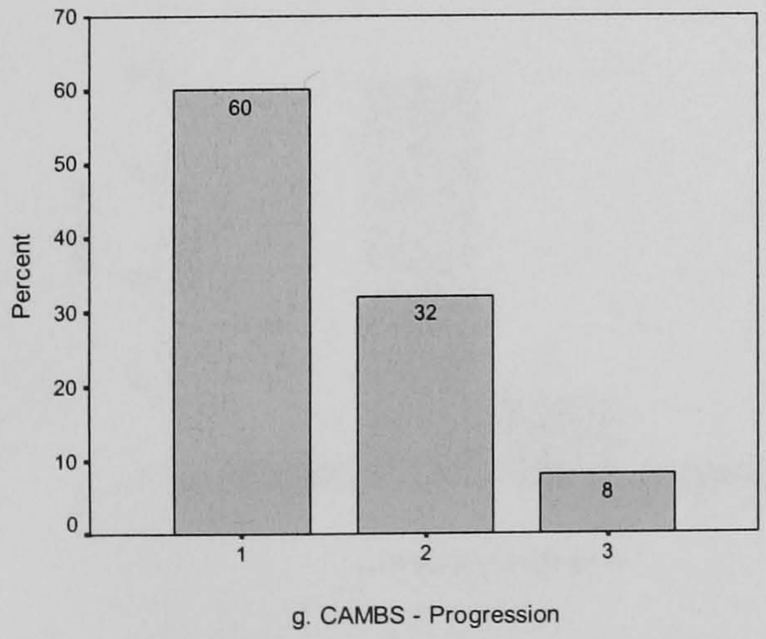
d. Ambulation Index



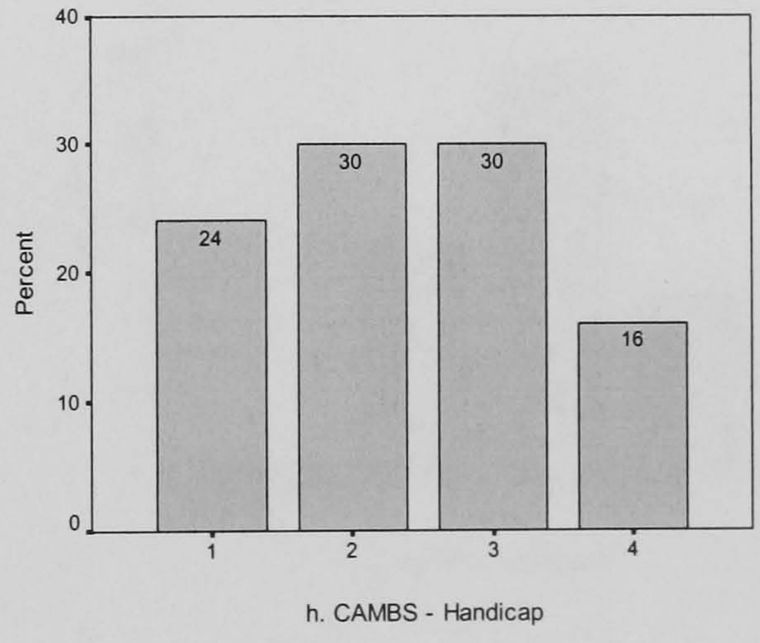
e. CAMBS - Disability



f. CAMBS - Relapse

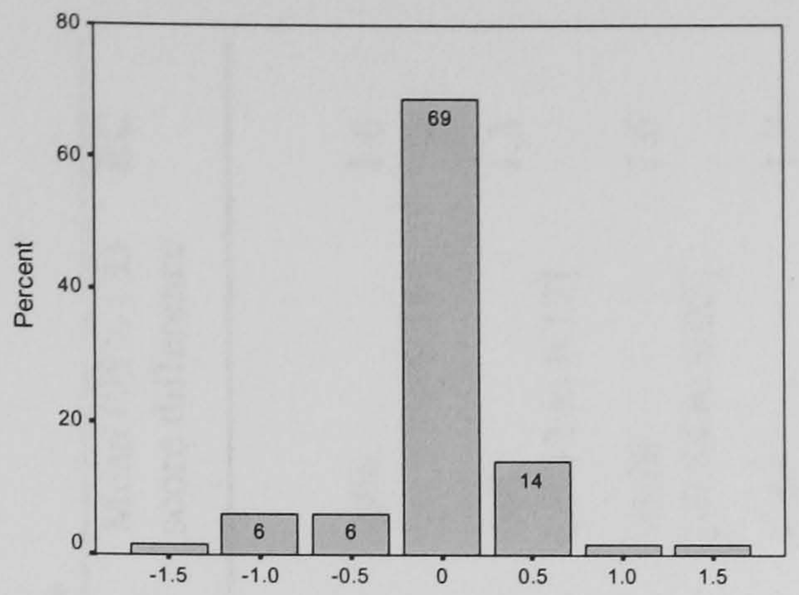


g. CAMBS - Progression

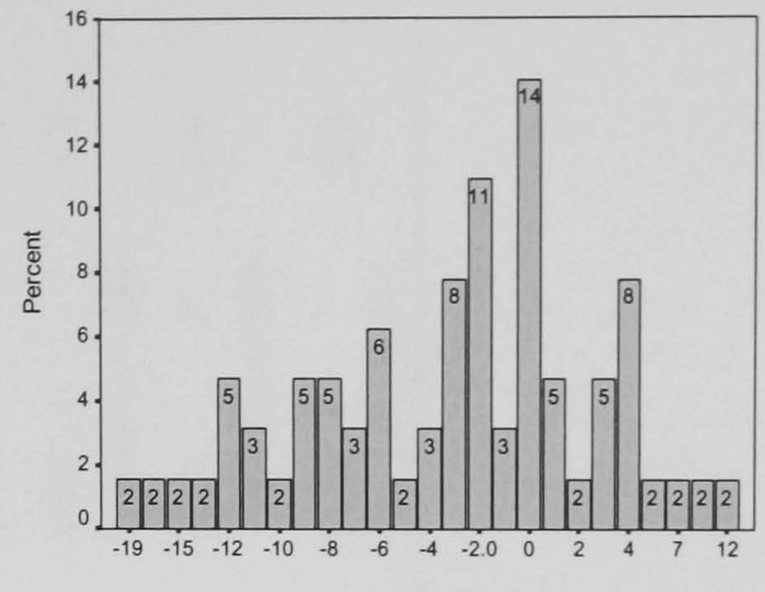


h. CAMBS - Handicap

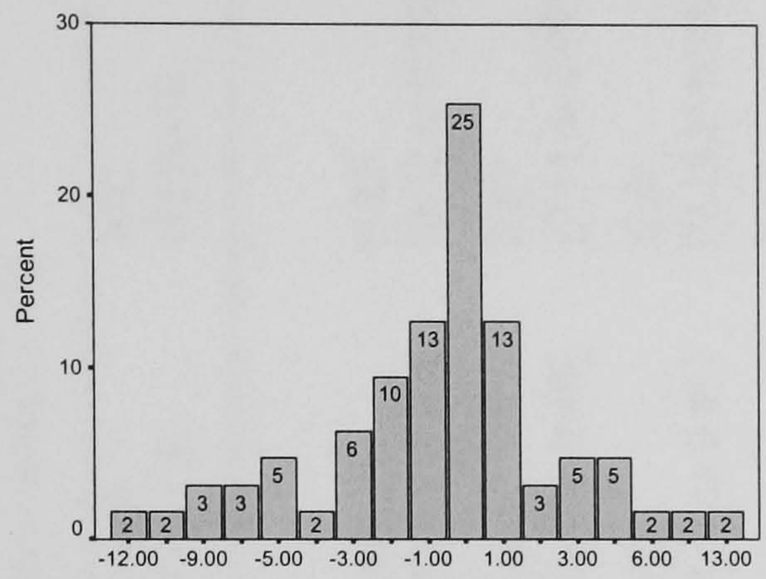
Figure 5.2 Score frequency distributions of the EDSS, SNRS, FIM, Ambulation Index (n = 64), and CAMBS (n = 50)



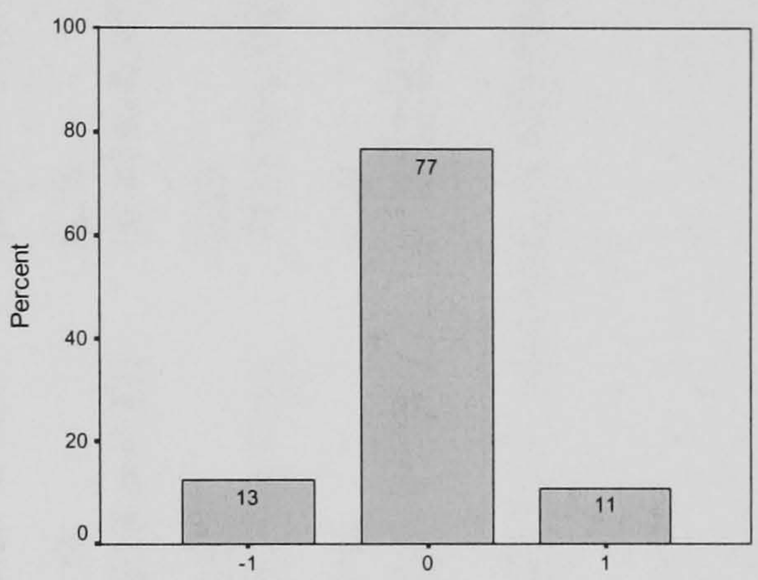
a. EDSS



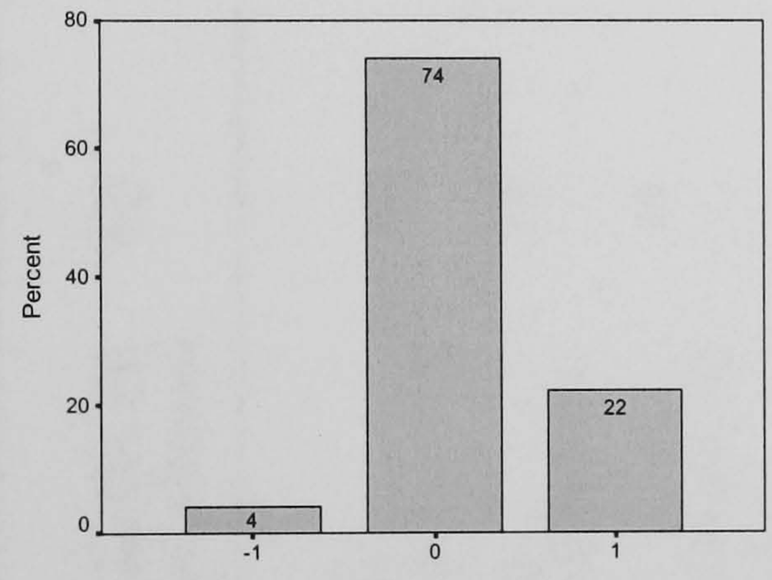
b. SNRS



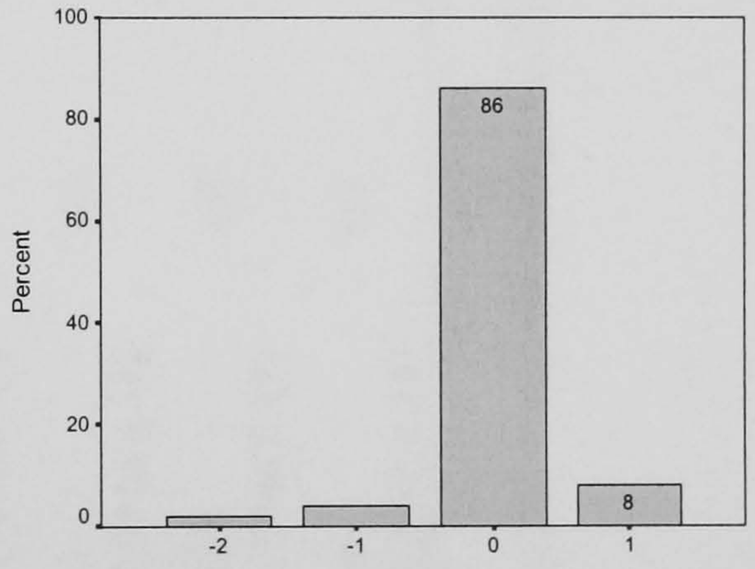
c. FIM



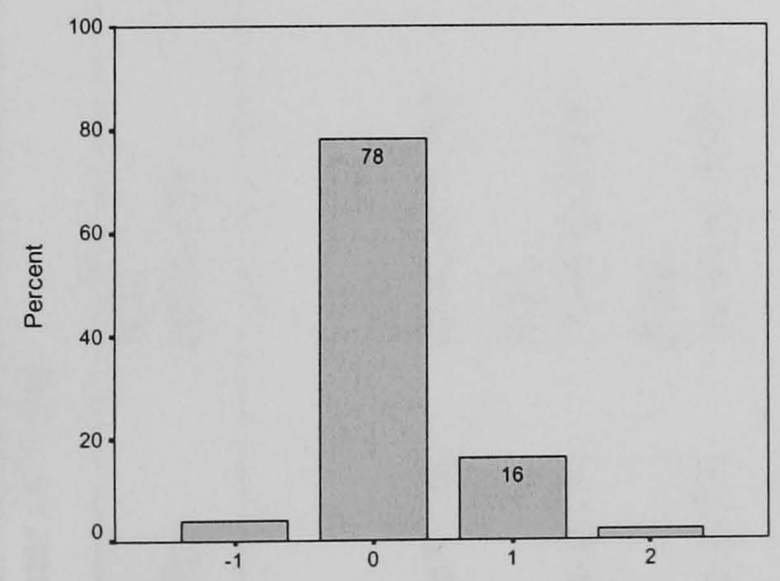
d. Ambulation Index



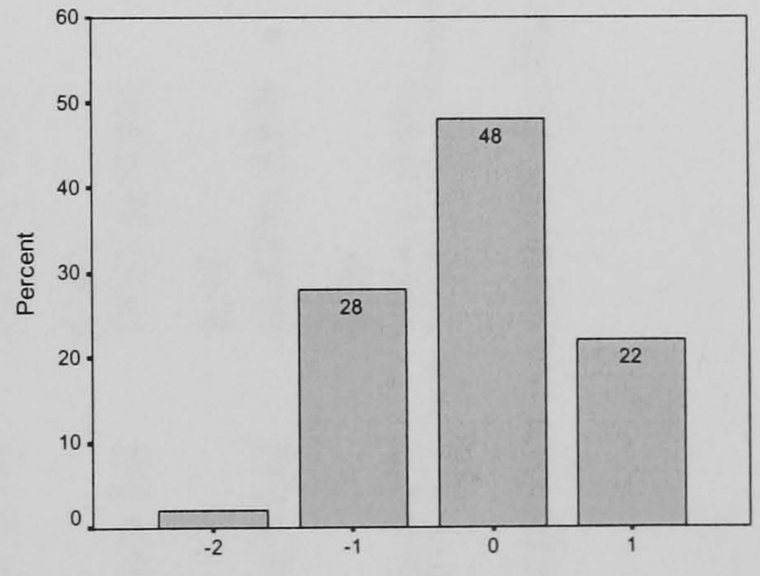
e. CAMBS - Disability



f. CAMBS - Relapse



g. CAMBS - Progression



h. CAMBS - Handicap

Figure 5.3 Inter-rater reliability: score differences between the two raters for the EDSS, SNRS, FIM, Ambulation Index (n = 64), and CAMBS (n = 50)

PAGE

MISSING

IN

ORIGINAL

Table 5.1 Reliability (inter-and intra-rater) of the Functional Systems and the EDSS

Scale	Inter-rater (n = 64)				Intra-rater (n = 35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Functional Systems								
Pyramidal	0.63 [0.50 to 0.76]	0.95 [0.80 to 0.99]	0.04 [- 0.10 to 0.19]	1.2	0.64 [0.46 to 0.82]	0.92 [0.72 to 0.99]	0.06 [-0.29 to 0.18]	1.6
Cerebellar	0.61 [0.47 to 0.75]	0.91 [0.68 to 0.99]	0.01 [-0.36 to 0.42]	1.4	0.66 [0.46 to 0.86]	0.67 [0.11 to 0.99]	001 [-0.12 to 0.12]	1.3
Brain stem	0.59 [0.41 to 0.72]	0.88 [0.68 to 0.99]	0 [- 0.19 to 0.19]	1.5	0.63 [0.38 to 0.88]	0.67 [0.15 to 0.99]	-0.26 [-0.54 to 0.02]	1.6
Bladder & Bowel	0.63 [0.49 to 0.77]	0.95 [0.82 to 0.99]	0.03 [- 0.12 to 0.19]	1.2	0.60 [0.40 to 0.79]	0.92 [0.53 to 0.99]	0.14 [-0.15 to 0.4.3]	1.7
Sensory	0.41 [0.28 to 0.54]	0.81 [0.42 to 0.99]	0.18 [- 0.13 to 0.47]	1.5	0.43 [0.23 to 0.63]	0.86 [0.55 to 0.99]	-0.17 [-0.45 to 0.11]	1.6
Mental	0.42 [0.27 to 0.56]	0.87 [0.57 to 0.99]	0.17 [- 0.03 to 0.37]	1.6	0.58 [0.34 to 0.82]	0.78 [0.40 to 0.99]	0.09 [-0.16 to 0.33]	1.4
Visual	0.67 [0.53 to 0.81]	0.95 [0.83 to 0.99]	-0.016 [- 0.19 to 0.17]	1.4	0.42 [0.20 to 0.64]	0.88 [0.51 to 0.99]	0.14 [-0.46 to 0.17]	1.8
EDSS	0.65 [0.53 to 0.77]	0.99 [0.96 to 0.99]	0.08 [- 0.20 to 0.10]	0.9	0.70 [0.53 to 0.87]	0.99 [0.95 to 0.99]	0.14 [-0.13 to 0.16]	0.8

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

Table 5.2 Reliability (inter-and intra-rater) of the SNRS

Items	Inter-rater (n =64)				intra-rater (n =35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Mentation and mood	0.62 [0.46 to 0.78]	0.87 [0.57 to 0.99]	0.16 [-0.26 to 0.57]	3.3	0.55 [0.28 to 0.82]	0.65 [0.10 to 0.99]	0.03 [-0.49 to 0.55]	2.9
Cranial nerves								
Visual acuity	0.65 [0.51 to 0.79]	0.93 [0.77 to 0.99]	-0.11 [-0.28 to 0.06]	1.4	0.75 [0.55 to 0.95]	0.89 [0.65 to 0.99]	-0.03 [-0.27 to 0.21]	1.4
Fields, discs, pupils	0.35 [0.18 to 0.52]	0.62 [0.22 to 0.99]	-0.67 [-0.98 to –0.36]	2.4	0.45 [0.10 to 0.80]	0.56 [0.05 to 0.99]	-0.03 [-0.41 to 0.34]	2.1
Eye movements	0.34 [0.21 to 0.47]	0.76 [0.31 to 0.99]	-0.50 [-0.81 to –0.46]	2.4	0.58 [0.29 to 0.86]	0.60 [0.08 to 0.99]	0.11 [-0.36 to 0.59]	2.2
Nystagmus	0.30 [0.20 to 0.40]	0.71 [0.20 to 0.99]	-0.50 [-0.85 to –0.15]	2.7	0.53 [0.33 to 0.86]	0.64 [0.09 to 0.99]	0.09 [-0.33 to 0.49]	2.4
Lower cranial nerves	0.61 [0.46 to 0.76]	0.92 [0.74 to 0.99]	-0.28 [-0.44 to –0.12]	1.2	0.51 [0.15 to 0.87]	0.75 [0.29 to 0.99]	0.23 [0.07 to 0.45]	1.3
Motor								
Right upper limb	0.66 [0.48 to 0.84]	0.93 [0.75 to 0.99]	-0.05 [-0.22 to 0.12]	1.1	0.49 [0.28 to 0.69]	0.65 [0.10 to 0.99]	0.11 [-0.23 to 0.46]	1.9
Left upper Limb	0.72 [0.54 to 0.89]	0.93 [0.76 to 0.99]	-0.08 [-0.24 to 0.08]	1.3	0.33 [0.00 to 0.66]	0.56 [0.04 to 0.99]	0.03 [-0.35 to 0.41]	2.2
Right lower limb	0.68 [0.55 to 0.81]	0.93 [0.68 to 0.99]	-0.33 [-0.54 to 0.11]	1.7	0.60 [0.41 to 0.79]	0.85 [0.54 to 0.99]	0.29 [-0.10 to 0.67]	2.2
Left lower Limb	0.62 [0.48 to 0.76]	0.91 [0.61 to 0.99]	-0.29 [-0.55 to –0.04]	1.9	0.62 [0.43 to 0.81]	0.84 [0.49 to 0.99]	0.37 [-0.05 to 0.79]	2.4
Tendon reflexes								
Upper limbs	0.43 [0.25 to 0.61]	0.89 [0.41 to 0.99]	0.23 [-0.02 to 0.58]	2.4	0.42 [0.14 to 0.70]	0.62 [0.12 to 0.99]	0 [-0.22 to 0.22]	1.3
Lower limbs	0.44 [0.28 to 0.59]	0.59 [0.11 to 0.99]	-0.28 [-0.64 to -0.07]	2.8	0.45 [0.24 to 0.66]	0.59 [0.05 to 0.99]	0.20 [-0.29 to 0.69]	2.8

Items	Inter-rater (n =64)				intra-rater (n =35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Babinski								
Right	0.72 [0.51 to 0.93]	0.72 [0.27 to 0.99]	0.05 [-0.15 to 0.18]	1.3	0.72 [0.48 to 0.97]	0.72 [0.23 to 0.99]	0.23 [-0.02 to 0.48]	0.8
Left	0.66 [0.43 to 0.89]	0.66 [0.18 to 0.99]	-0.02 [-0.19 to 0.16]	1.4	0.52 [0.20 to 0.84]	0.52 [0.07 to 0.99]	0.23 [-0.05 to 0.51]	1.6
Sensory								
Right Upper limb	0.48 [0.25 to 0.71]	0.67 [014 to 0.99]	-0.11 [-0.27 to 0.05]	1.3	0.39 [0.02 to 0.76]	0.81 [0.42 to 0.99]	0.06 [-0.09 to 0.20]	0.8
Left Upper Limb	0.41 [0.20 to 0.62]	0.67 [0.12 to 0.99]	-0.09 [-0.25 to 0.07]	1.3	0.49 [0.18 to 0.79]	0.74 [0.28 to 0.99]	0.17 [0.02 to 0.33]	09
Right Lower limb	0.48 [0.33 to 0.63]	0.81 [0.43 to 0.99]	-0.03 [-0.20 to 0.14]	1.4	0.43 [0.20 to 0.66]	0.83 [0.49 to 0.99]	0.08 [-0.11 to 0.28]	1.1
Left Lower Limb	0.55 [0.40 to 0.69]	0.88 [0.60 to 0.99]	0.05 [-0.02 to 0.18]	1.1	0.73 [0.54 to 0.92]	0.92 [0.72 to 0.99]	0.03 [-0.10 to 0.16]	0.8
Cerebellar								
Upper limbs	0.52 [0.36 to 0.68]	0.83 [0.46 to 0.99]	-0.29 [-0.59 to –0.09]	2.3	0.60 [0.38 to 0.82]	0.85 [0.52 to 0.99]	-0.14 [-0.46 to 0.17]	1.8
Lower limb	0.69 [0.55 to 0.83]	0.86 [0.55 to 0.99]	0 [-0.29 to 0.29]	2.1	0.71 [0.51 to 0.91]	0.79 [0.39 to 0.99]	-0.34 [-0.81 to 0.13]	2.6
Gait, trunk & balance	0.72 [0.59 to 0.85]	0.93 [0.74 to 0.99]	0.06 [-0.30 to 0.43]	2.9	0.61 [0.40 to 0.82]	0.71 [0.21 to 0.99]	0.06 [-0.77 to 0.88]	4.7
Bladder, bowel & sexual	0.44 [0.29 to 0.59]	0.54 [0.08 to 0.99]	1.20 [0.61 to 0.179]	4.6	0.47 [0.23 to 0.71]	0.61 [0.10 to 0.99]	-0.34 [-1.15 to 0.46]	4.6
Total SNRS score	0.12 [-0.07 to 0.32]	0.97 [0.91 to 0.99]	-3.08 [-4.62 to –1.53]	12.1	0.04 [-0.16 to 0.20]	0.94 [0.79 to 0.99]	1.40 [-0.75 to 3.54]	12.2

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

Table 5.3 Reliability (inter-and intra-rater) of the FIM

Items	Inter-rater (n =64)				Intra-rater (n = 35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Self-care								
A) Eating	0.61 [0.45 to 0.77]	0.91 [0.71 to 0.99]	-0.127 [-0.33 to 0.07]	1.6	0.68 [0.41 to 0.95]	0.81 [0.42 to 0.99]	0.2 [-0.02 to 0.41]	1.2
B) Grooming	0.72 [0.56 to 0.88]	0.95 [0.82 to 0.99]	0 [-0.16 to 0.16]	1.3	0.66 [0.36 to 0.96]	0.73 [0.20 to 0.99]	0 [-0.14 to 0.14]	0.8
C) Bathing	0.56 [0.39 to 0.73]	0.97 [0.89 to 0.99]	0.04 [-0.09 to 0.19]	1.1	0.77 [0.55 to 0.99]	0.94 [0.81 to 0.99]	0 [-0.09 to 0.09]	0.5
D) Dressing-upper body	0.77 [0.62 to 0.92]	0.95 [0.82 to 0.99]	-0.03 [-0.02 to 0.14]	1.4	0.78 [0.59 to 0.97]	0.97 [0.91 to 0.99]	-0.06 [-0.14 to 0.02]	0.5
E) Dressing-lower body	0.66 [0.52 to 0.80]	0.97 [0.89 to 0.99]	-0.05 [-0.18 to 0.08]	1.0	0.55 [0.30 to 0.80]	0.96 [0.85 to 0.99]	-0.03 [-0.13 to 0.07]	0.6
F) Toileting	0.74 [0.58 to 0.89]	0.98 [0.93 to 0.99]	0.05 [-0.07 to 0.16]	0.9	0.77 [0.56 to 0.98]	0.86 [0.55 to 0.99]	-0.06 [-0.14 to 0.02]	0. 5
Sphincter control								
G) Bladder	0.74 [0.61 to 0.87]	0.89 [0.65 to 0.99]	0.03 [-0.20 to 0.03]	2.1	0.57 [0.36 to 0.78]	0.60 [0.10 to 0.99]	-0.17 [0.45 to 0.12]	1.6
H) Bowel	0.74 [0.57 to 0.91]	0.94 [0.79 to 0.99]	-0.02 [-0.04 to –0.02]	1.3	0.77 [0.57 to 0.97]	0.77 [0.33 to 0.99]	0.06 [-0.02 to 0.14]	0.5
Mobility - Transfer								
I) Transfer bed/chair	0.77 [0.65 to 0.89]	0.98 [0.94 to 0.99]	-0.06 [-0.18 to 0.02]	0.8	0.66 [0.40 to 0.92]	0.72 [0.31 to 0.99]	-0.06 [-0.22 to 0.01]	0.9
J) Transfer toilet	0.30 [0.16 to 0.44]	0.95 [0.86 to 0.99]	0.05 [-0.06 to 0.16]	0.9	0.60 [0.34 to 0.87]	0.65 [0.11 to 0.99]	0.06 [-0.09 to 0.02]	0.8
K) Transfer tub/shower	0.87 [0.77 to 0.97]	0.99 [0.98 to 0.99]	-0.05 [-0.12 to 0.02]	0.6	0.75 [0.55 to 0.95]	0.88 [0.60 to 0.99]	-0.03 [-0.24 to 0.18]	1.2

Items	Inter-rater (n =64)				Intra-rater (n = 35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Mobility - locomotion								
L) Walking	0.64 [0.51 to 0.77]	0.93 [0.77 to 0.99]	-0.16 [-0.38 to 0.06]	1.7	0.90 [0.80 to 0.99]	0.99 [0.96 to 0.99]	0 [-0.08 to 0.08]	0.5
M) Stairs	0.88 [0.77 to 0.99]	0.99 [0.97 to 0.99]	0.08 [0.01 to 0.15]	0.5	0.75 [0.55 to 0.95]	0.80 [0.49 to 0.99]	0 [-0.39 to 0.39]	2.2
Communication								
N) Comprehension	0.26 [0.13 to 0.39]	0.56 [0.31 to 0.99]	-0.23 [-0.58 to 0.10]	2.7	0.66 [0.44 to 0.88]	0.99 [0.81 to 0.99]	-0.03 [-0.03 to 0.09]	0.3
O) Expression	0.40 [0.25 to 0.55]	0.76 [0.31 to 0.99]	0.03 [-0.21 to 0.28]	1.9	1 [NA]	1 [N/A]	0 [NA]	0
Social cognition								
P) Social interaction	0.55 [0.42 to 0.68]	0.92 [0.71 to 0.99]	0.02 [-0.14 to 0.17]	1.2	1 [NA]	1 [N/A]	0 [NA]	0
Q) Problem solving	0.55 [0.40 to 0.70]	0.92 [0.72 to 0.99]	0.03 [-0.14 to 0.21]	1.4	0.66 [0.46 to 0.86]	0.97 [0.89 to 0.99]	0.03 [-0.03 to 0.09]	0.3
R) Memory	0.49 [0.33 to 0.65]	0.94 [0.76 to 0.99]	0.02 [-0.14 to 0.17]	1.2	0.65 [0.41 to 0.89]	0.80 [0.40 to 0.99]	0.03 [-0.16 to 0.23]	1.1
Total FIM score	0.21 [0.01 to 0.41]	0.99 [0.97 to 0.99]	-0.66 [-1.71 to 0.37]	8.1	0.28 [0.04 to 0.52]	0.94 [0.80 to 0.99]	-0.06 [-1.13 to 1.102]	6.1

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

Inter-rater reliability agreement on the different domains was variable with kappa coefficients ranging between 0.45 and 0.69 (moderate to substantial), intraclass correlation coefficients ranging between 0.61 and 0.88 (substantial to almost perfect), and repeatability coefficients ranging between 0.9 and 1 points (Table 5.4). The largest score differences between the two raters were 1 point for the disability domain, 2 points for the relapse domain, 2 points for the progression domain, and 2 points for the handicap domain (Figure 5.3e – 5.3h).

F). Raters' bias

With the exception of the cranial nerves item, bladder, bowel and sexual item, and some of the motor and the cerebellar items of the SNRS, the mean score differences between the two raters were generally small with narrow 95% confidence intervals which included the "0" value indicating the absence of raters' bias.

5.3.2 Intra-rater reliability

Thirty-five patients had remained stable between two visits on at least one occasion during the 9 months follow up period. To avoid introducing any statistical bias, only one pair of assessments (the first) per patient was included in the final analysis. This cohort consisted of 20 women and 15 men with a median age of 38 years (range 24-51 years), and median disease duration of 11 years (2-17 years). To compensate for the design of the relapse and the progression domains of the CAMBS (which have been devised to assess disease stability over a time longer than three months), intra-rater reliability of these two domains was assessed in a sub-group of 23 patients after excluding all patients who had any relapses during the 9 months before the first assessment (9 patients) or between the first and the second assessments (3 patients - all had mild relapses which recovered completely).

Table 5.4 Reliability (inter-and intra-rater) of the CAMBS, the Ambulation Index, and the Barthel Index

Scale	Inter-rater (n = 50)				Intra-rater (n = 35)			
	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Ambulation Index *	0.73 [0.61 to 0.85]	0.96 [0.95 to 0.99]	-0.02 [-0.14 to 0.11]	1	0.59 [0.41 to 0.77]	0.93 [0.76 to 0.99]	0.25 [-0.01 to 0.51]	1.5
CAMBS								
Disability	0.56 [0.37 to 0.75]	0.80 [0.40 to 0.99]	0.18 [0.04 to 0.32]	1	0.67 [0.50 to 0.87]	0.85 [0.54 to 0.99]	-0.06 [-0.22 to 0.11]	0.8
Relapse	0.69 [0.51 to 0.87]	0.79 [0.38 to 0.99]	0 [-0.13 to 0.13]	0.9	0.80 [0.60 to 0.99]	0.85 [0.28 to 0.99]	0.06 [-0.09 to 0.020]	0.8
Progression	0.54 [0.43 to 0.66]	0.61 [0.23 to 0.99]	0.16 [0.0 to 0.30]	1	0.58 [0.34 to 0.82]	0.71 [0.21 to 0.99]	0.06 [-0.09 to 0.20]	0.9
Handicap	0.45 [0.33 to 0.57]	0.88 [0.61 to 0.99]	0.04 [-0.10 to 0.18]	1	0.57 [0.38 to 0.76]	0.72 [0.23 to 0.99]	-0.03 [-0.27 to 0.21]	1.4
Barthel Index	N/A	N/A	N/A	N/A	0.75 [0.56 to 0.94]	0.98 [0.91 to 0.99]	-0.03 [-0.27 to 0.21]	1.1

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals; * n = 64 patients

A). EDSS

The median (range) scores of the two assessments were identical at 4.5 (0-7.5). Intra-rater agreement on the different Functional System scores was variable, with kappa coefficients ranging between 0.42 and 0.66 (fair to substantial), intraclass correlation coefficients ranging between 0.67 and 0.92 (substantial to almost perfect), and repeatability coefficients ranging between 1.3 and 1.8 points (Table 5.1). The largest score differences between the two assessments were 2 points for the pyramidal, sensory, bladder and bowel, and mental Functional Systems, and 3 points for the cerebellar, brain stem, and visual Functional Systems. Intra-rater agreement on the EDSS scores was 63%, 89%, and 100% when agreement was defined as no difference, a difference of ≤ 0.5 point (one EDSS step), and ≤ 1.0 point (two 0.5 EDSS steps) respectively (Figure 5.4a), with a repeatability coefficient of 0.8 point, a kappa coefficient of 0.7 (substantial), and an intraclass correlation coefficient of 0.99 (almost perfect) (Table 5.1).

B). SNRS

The median (range) scores for the two assessments were very similar at 73 (33-98) and 71 (34-98). Intra-rater agreement on the different scale items was variable with kappa coefficients ranging between 0.33 and 0.75 (fair to substantial), intraclass correlation coefficients ranging between 0.52 and 0.92 (moderate to almost perfect), and repeatability coefficients ranging between 0.8 and 4.7 points (Table 5.2). Intra-rater agreement on the sum scores was 6%, 67%, 76%, and 100% when agreement was defined as no difference, a difference of ≤ 5 points, ≤ 10 points, and ≤ 14 points respectively (Figure 5.4b), with a repeatability coefficient of 12.2 points, and an intraclass correlation coefficient of 0.94 (almost perfect) (Table 5.2).

C). FIM

The median (range) scores of the two assessments were identical at 123 (90-126). Intra-rater agreement on the different scale items was variable with kappa coefficients ranging between 0.55 and 1 (moderate to perfect), intraclass correlation coefficients ranging between 0.60 and 1 (substantial to perfect), and repeatability coefficients ranging between 0 and 2.2 points (Table 5.3). Intra-rater agreement on

the sum scores was 37%, 92%, and 100% when agreement was defined as no difference, a difference of ≤ 5 points, and ≤ 9 points respectively (Figure 5.4c), with a repeatability coefficients of 6.1 points, and an intraclass correlation coefficient of 0.94 (almost perfect) (Table 5.3).

D). AI

The median (range) scores of the two assessments were identical at 2 (0-8). Intra-rater agreement was 66%, 94%, 97%, and 100% when agreement was defined as no difference, a difference ≤ 1 points, ≤ 2 points, and ≤ 3 points respectively (Figure 5.4d), with a repeatability coefficient of 1.5 points, a kappa coefficient of 0.59 (moderate), and an intraclass correlation coefficient of 0.93 (almost perfect) (Table 5.4).

E). CAMBS

The median (range) scores of the two assessments for the scale's four domains were identical: disability 2 (1-4), relapse 1 (1-3), progression 1 (1-3), and handicap 2 (1-4). Intra-rater agreement on the different domains was very high with kappa coefficients ranging between 0.58 and 0.80 (moderate to substantial), intraclass correlation coefficients ranging between 0.71 and 0.85 (substantial to almost perfect), and repeatability coefficients ranging between 0.8 and 1.4 points (Table 5.4). The largest score differences between the two assessments were 1 point for the disability domain, 2 points for the relapse domain, 1 point for the progression domain, and 2 points for the handicap domain (Figure 5.4e–5.4h).

F). Raters' bias

With the exception of the lower cranial nerves item of the SNRS, the mean score differences between the two raters were generally small with narrow 95% confidence intervals which included the "0" value indicating the absence of raters' bias.

5.3.3 Internal consistency and factor analysis

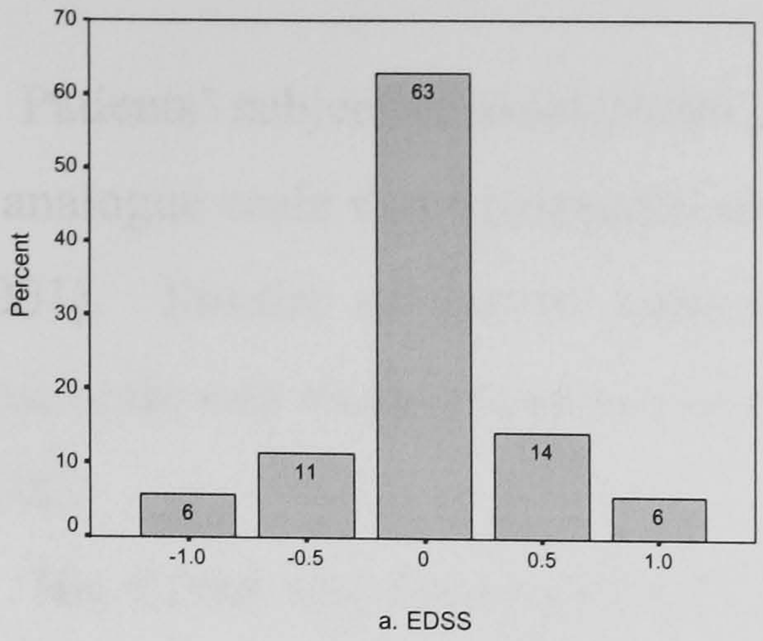
Internal consistency and factor analysis of the two multidimensional scales assessed in this study (SNRS and FIM) were evaluated using the inter-rater reliability data set. Internal consistency was very high with Cronbach's alpha of 0.92

for the SNRS and 0.98 for the FIM. Factor analysis of SNRS suggested a five factor solution which accounted for 79.3% of the total variance (cumulative percentage of 52.8%, 63.0%, 69.7%, 74.8%, and 79.3%; eigenvalues of 11.6, 2.2, 1.5, 1.1, 1 respectively). The first factor of the rotated matrix (cerebellar factor) correlated with the "upper and lower limb cerebellar" (the latter also correlated with the fourth factor), "eye movements", "lower cranial nerves", and "nystagmus" items. The second factor (cerebral / visual / upper limb motor factor) correlated with the "mentation and mood", "visual acuity", "fields/discs/pupils", "upper limb motor", and "reflexes" items. The third factor (sensory factor) correlated with the "upper and lower limb sensory" (the latter also correlated with the fourth factor) items. The fourth factor (lower limb / spinal factor) correlated with the "lower limb motor", "lower limb cerebellar", "gait", and "bladder & bowel & sexual function" items. The fifth factor (Babinski factor) correlated with the "Babinski reflex" item only.

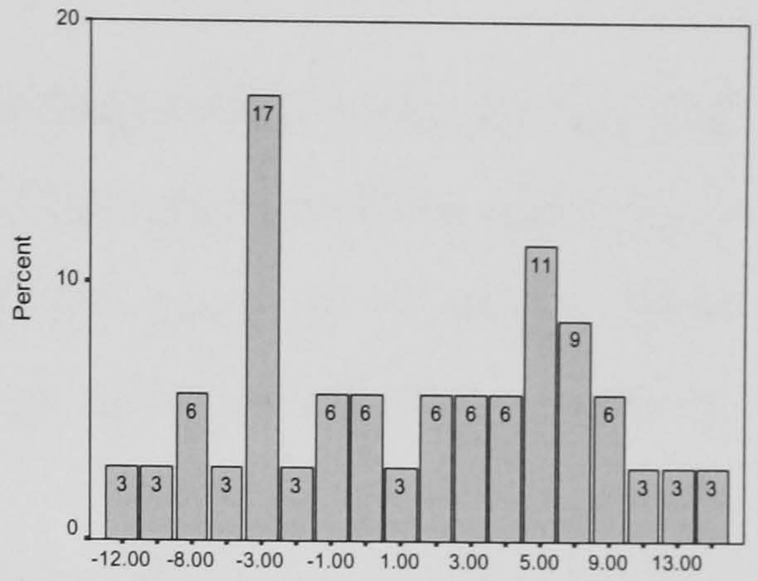
Factor analysis of the FIM suggested a two factor solution which accounted for 89.4% of the total variance (cumulative percentage of 83% and 89.4%; eigenvalues of 14.9 and 1.2 respectively). The first factor of the rotated matrix (motor factor) correlated with the "motor" items of the scale (items A to M), and the second factor (cognitive factor) correlated with the "communication" and "social cognition" items (items N to R).

5.3.4 Responsiveness

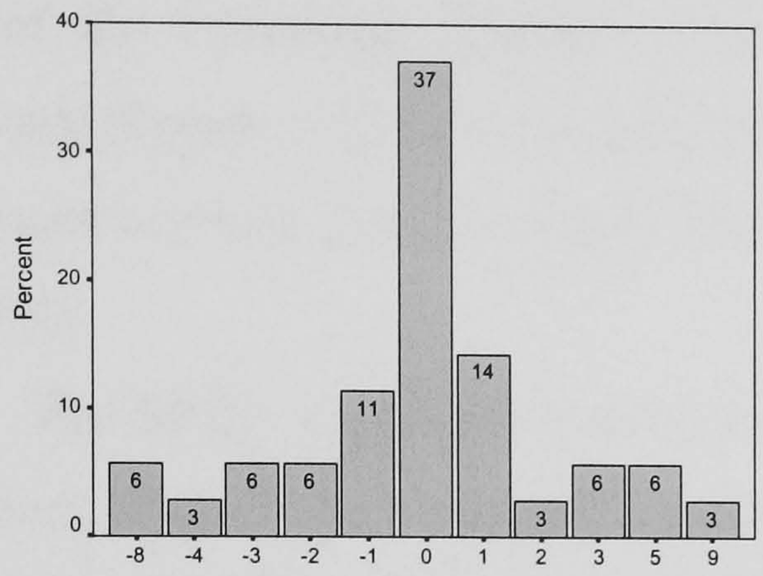
Of the 50 patients assessed, 25 were found to have changed on at least one occasion during the 9 months follow up period. This group consisted of 20 women and 5 men with a median age of 36 years (range 24-51), median EDSS of 5.5 (range 0–7.5), and a median disease duration of 10 years (range 2-22). To avoid introducing any statistical bias, only one pair of assessments (the first) per patient was included in the final analysis. The order of assessment in each pair (15 patients worsened, and 10 improved) was latter re-arranged so as to make all the changes of one direction (stable or improved to worsened).



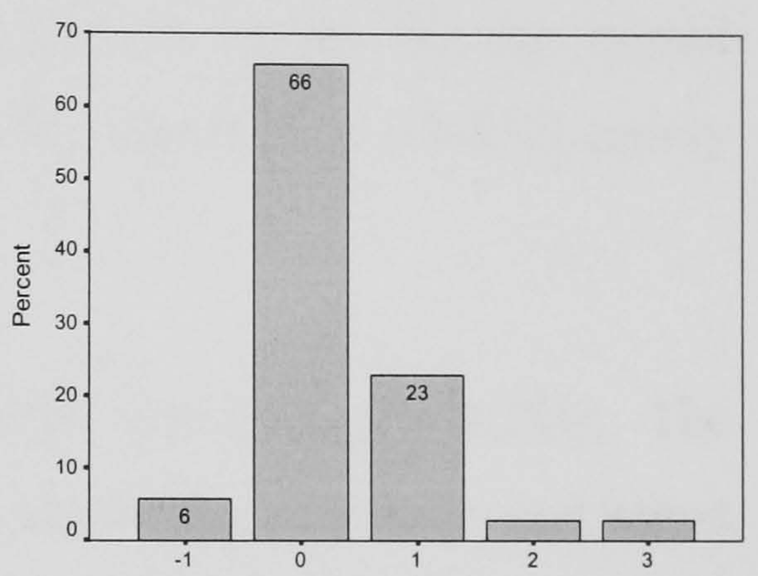
a. EDSS



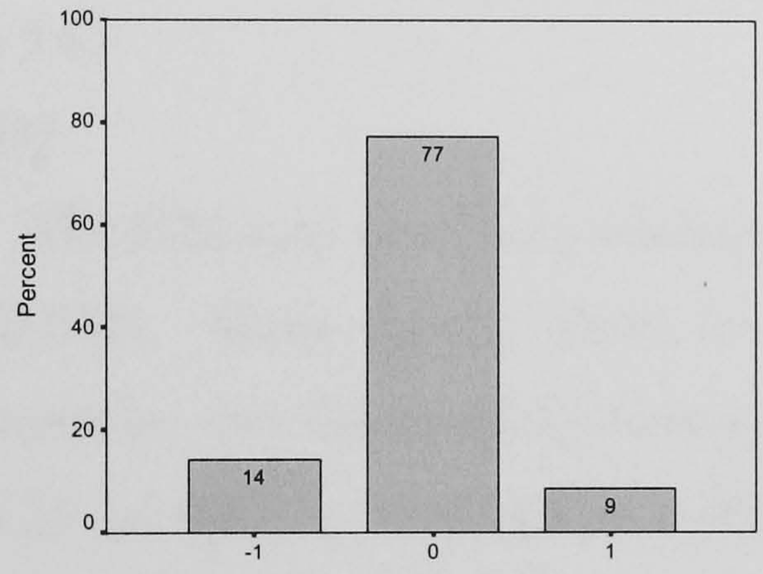
b. SNRS



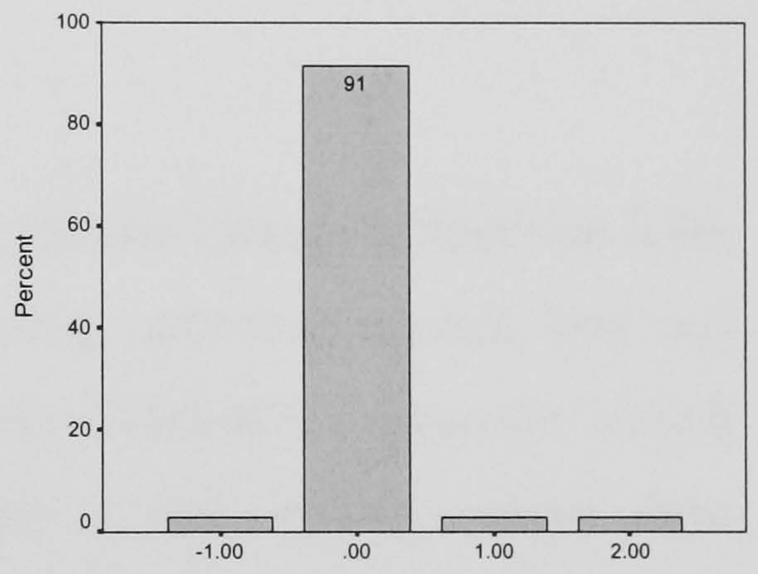
c. FIM



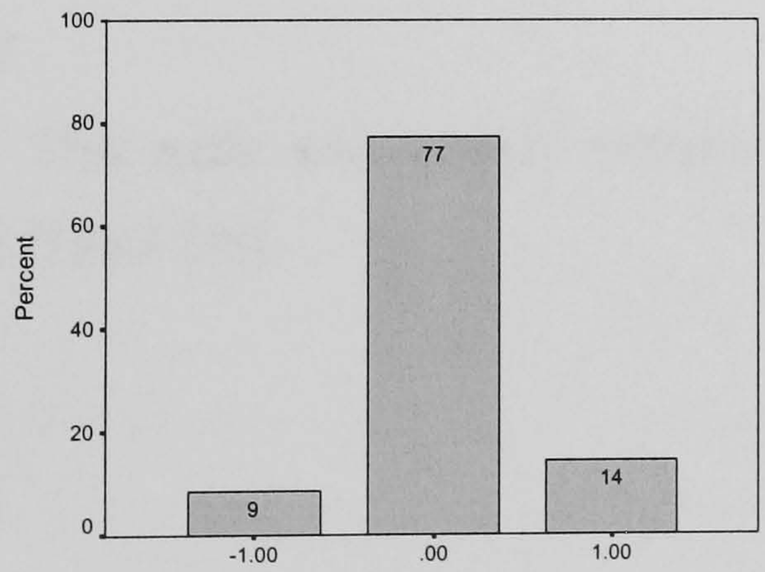
d. Ambulation Index



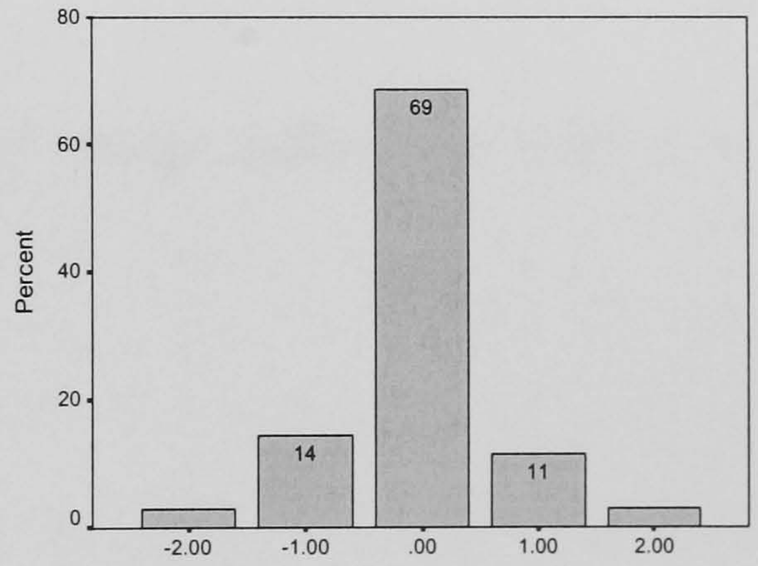
e. CAMBS - Disability



f. CAMBS - Relapse *



g. CAMBS - Progression *



h. CAMBS - Handicap

Figure 5.4 Intra-rater reliability: score differences between the two assessments for the EDSS, SNRS, FIM, Ambulation Index, and CAMBS (n = 35)

(* relapse & progression: n = 23)

Patients' subjective assessments of their own health status using the EuroQol visual analogue scale was moderately sensitive to clinical change (effect size 0.55, p = <0.001). Similar subjective assessment by myself using the EuroQol visual analogue scale was weakly sensitive to clinical change (effect size 0.36, p = <0.001).

A). EDSS

The EDSS was not sensitive to clinical change (effect size 0.11, p = 0.051). Most of the Functional Systems were also unresponsive except for the mental Functional System which was weakly responsive (effect size 0.38, p = 0.012) mainly on account of mood changes (Table 5.5).

B). SNRS

The SNRS sum score was unresponsive (effect size 0.17, p = 0.253). The individual scale items were also unresponsive except for the mentation and mood item which was weakly sensitive to clinical change (effect size 0.36, p = 0.043) (Table 5.6).

C). FIM

The FIM sum score was weakly sensitive to clinical change (effect size 0.46, p = <0.001). Many 'motor' items (eating, grooming, sphincter control, bed and toilet transfer, and locomotion) were also weakly to moderately responsive (effect size 0.25 to 0.67, p = 0.044 to 0.039), but none of the cognitive items were responsive (Table 5.7).

D). AI

This scale was weakly sensitive to clinical change (effect size 0.20, p = 0.039) (Table 5.8).

Table 5.5 Responsiveness of the Functional Systems and the EDSS (n = 25)

Scale	Time 1 Median (range)	Time 2 Median (range)	P *	Effect size
Functional Systems				
Pyramidal	2 [0 to 5]	3 [0 to 5]	0.160	0.13
Cerebellar	1 [0 to 5]	2 [0 to 3]	0.766	- 0.06
Brain stem	0 [0 to 4]	0 [0 to 3]	0.683	0.06
Bladder & Bowel	2 [0 to 4]	2 [0 to 4]	0.470	0.13
Sensory	0 [0 to 6]	1 [0 to 6]	0.210	0.14
Mental	0 [0 to 2]	1 [0 to 3]	0.012	0.38
Visual	0 [0 to 5]	1 [0 to 5]	0.386	0.09
EDSS	5.5 [0 to 6.5]	6 [0 to 7.5]	0.051	0.11

* Wilcoxon Signed Ranks test

E). CAMBS

The relapse and progression domains were moderately sensitive to clinical change (effect size 0.67, p = 0.001 and 0.78, p = <0.001 respectively), whereas the disability domain was only weakly responsive (effect size 0.39, p = 0.008), and the handicap domain was unresponsive (effect size 0.14, p = 0.206) (Table 5.8).

To assess the responsiveness of the five scales at different levels of disease severity, the patients were categorised into one of three levels of disease severity according to their baseline EDSS scores: mild (EDSS 0.0-4.5), moderate (EDSS 5.0-6.0), and severe (EDSS 6.5-7.5). Sub-group analysis showed the responsiveness of these scales within each band of disease severity to be similar to the results of the whole group.

Table 5.6 Responsiveness of the Scripps Neurological Rating Scale ($n = 25$)

	Time 1 Median (range)	Time 2 Median (range)	*P* *	Effect size
Mentation and mood	10 [4 to 10]	7 [6 to 10]	0.043	0.36
Cranial nerves				
Visual acuity	5 [1 to 5]	5 [1 to 5]	0.414	0.14
Fields, discs, pupils	4 [2 to 4]	4 [2 to 6]	1	0
Eye movements	5 [0 to 5]	5 [0 to 5]	0.071	0.16
Nystagmus	5 [0 to 5]	5 [0 to 5]	0.461	0.17
Lower cranial nerves	5 [0 to 5]	5 [1 to 5]	0.396	0.1
Motor				
Right upper limb	5 [0 to 5]	5[0 to 5]	0.317	0.15
Left upper limb	5 [0 to 5]	5 [0 to 5]	1	0
Right lower limb	3 [0 to 5]	1 [0 to 5]	0.368	0.12
Left lower limb	3 [0 to 5]	3 [0 to 5]	0.831	0.04
Deep tendon reflexes				
Upper limbs	4 [1 to 4]	4 [1 to 4]	0.792	0.06
Lower limbs	1 [0 to 4]	3 [0 to 4]	0.455	0.12
Babinski				
Right	0 [0 to 2]	0 [0 to 2]	0.317	0.09
Left	0 [0 to 2]	0 [0 to 2]	0.157	0.17
Sensory				
Right Upper limb	3 [0 to 3]	3 [0 to 3]	0.564	- 0.09
Left Upper limb	3 [0 to 3]	3 [0 to 3]	0.414	- 0.05
Right Lower limb	3 [0 to 3]	2 [0 to 3]	0.132	0.21
Left Lower limb	3 [0 to 3]	3 [0 to 3]	0.317	0.09
Cerebellar				
Upper limb	5 [0 to 5]	5 [0 to 5]	0.317	- 0.13
Lower limb	5 [0 to 5]	3 [0 to 5]	0.546	0.09
Gait, trunk & balance	7 [0 to 10]	7 [0 to 10]	0.161	0.19
Bladder, bowel & sexual	-3 [-10 to 0]	-3 [-10 to 0]	0.951	0.06
Total SNRS score	70 [34 to 98]	66 [29 to 98]	0.253	0.17

* Wilcoxon Signed Ranks test

Table 5.7 Responsiveness of the Functional Independence Measure (n = 25)

	Time 1 Median (range)	Time 2 Median (range)	P *	Effect size
Self-care				
A) Eating	7 [3 to 7]	7 [3 to 7]	0.059	0.28
B) Grooming	7 [6 to 7]	7 [4 to 7]	0.015	0.39
C) Bathing	7 [3 to 7]	7 [3 to 7]	0.393	0.07
D) Dressing-upper body	7 [3 to 7]	7 [3 to 7]	0.593	0.07
E) Dressing-lower body	7 [3 to 7]	7 [3 to 7]	0.739	0.09
F) Toileting	7 [5 to -7]	7 [5 to 7]	0.655	0.07
Sphincter control				
G) Bladder	6 [1 to 7]	6 [1 to 7]	0.008	0.54
H) Bowel	7 [1 to 7]	7 [1 to 7]	0.48	-0.19
Mobility - Transfer				
I) Transfer bed/chair	7 [4 to 7]	7 [2 to 7]	0.039	0.67
J) Transfer toilet	7 [6 to 7]	7 [6 to 7]	0.048	0.25
K) Transfer tub/shower	6 [3 to 7]	6 [3 to 7]	0.832	0.13
Mobility - locomotion				
L) Walking	6 [2 to 7]	6 [2 to 7]	0.041	0.48
M) Stairs	6 [2 to 7]	6 [1 to 7]	0.038	0.66
Communication				
N) Comprehension	7 [4 to 7]	7 [4 to 7]	0.18	0
O) Expression	7 [6 to 7]	7 [4 to 7]	0.317	0
Social cognition				
P) Social interaction	7 [6 to 7]	7 [6 to 7]	1	0
Q) Problem solving	7 [3 to 7]	7 [3 to 7]	0.414	0.15
R) Memory	7 [3 to 7]	7 [3 to 7]	0.034	0.19
Total FIM score	123 [91 to 126]	120 [90 to 126]	<0.001	0.46

* Wilcoxon Signed Ranks test

Table 5.8 Responsiveness of the CAMBS, the Ambulation index, and the Barthel index (n = 25)

	Time 1 Median (range)	Time 2 Median (range)	P *	Effect size
Ambulation index	2 [0 to 6]	2 [0 to 9]	0.039	0.20
CAMBS				
Disability	2 [1 to 4]	2 [1 to 5]	0.008	0.39
Relapse	1 [1 to 2]	2 [1 to 4]	0.001	0.67
Progression	1 [1 to 2]	2 [1 to 3]	<0.001	0.78
Handicap	2 [1 to 4]	2 [1 to 5]	0.206	0.14
Barthel index	20 [10 to 20]	20 [10-20]	0.042	0.25

* Wilcoxon Signed Ranks test

5.3.5 Validity

Convergent validity was assessed on the same cohort of 50 patients who took part in the intra-rater reliability and responsiveness study. The group consisted of 31 women and 19 men, with a median age of 36 years (range 24–51), median EDSS score of 4.5 (range 0-7.5), and a median disease duration of 12 years (range 2-17).

A). Convergent and discriminant validity

The Barthel index correlated highly with the FIM (r = 0.88), and moderately with the EDSS (r = -0.74), SNRS (r = 0.69), AI (r = -0.72), and the disability and the handicap domains of the CAMBS (r = - 0.69, and -0.61 respectively) (Table 5.9). In comparison, the London Handicap Scale correlated moderately with the EDSS (r = -0.69), SNRS (r = 0.71), AI (r = -0.72), and the disability and the handicap domains of the CAMBS (r = -0.59 and -0.65 respectively), and weakly with the FIM (r = 0.43) (Table 5.9). The physical functioning item of the SF-36 correlated highly with the EDSS (r = -0.82), SNRS (r = 0.82), FIM (r = 0.88), AI (r = -0.87), and moderately with the disability and the handicap domains of the CAMBS (r = -0.71 and -0.65 respectively) (Table 5.9).

Table 5.9 Correlation between the SF-36, the London Handicap scale, the Barthel index and the EDSS, the SNRS, the Ambulation index, and the CAMBS

	EDSS	SNRS	*FIM*	AI	CAMBS - Disability	CAMBS - Handicap
Barthel Index	-0.74 *	0.69 *	0.88 *	-0.72 *	-0.69 *	-0.61 *
London handicap scale	-0.69 *	0.71 *	0.43 *	-0.72 *	-0.59 *	-0.65 *
EuroQol VAS	-0.69 *	0.67 *	0.69 *	-0.73 *	-0.55 *	-0.81 *
SF 36:						
Physical functioning	-0.82 *	0.82 *	0.88 *	- 0.87 *	-0.71 *	-0.65 *
Physical role limitation	- 0.50 *	0.46 **	0.36 ***	-0.52 *	-0.34	-0.54 *
Emotional role limitation	-0.11	0.19	0.15	-0.07	-0.09	-0.22
Social functioning	-0.47 *	0.37 * *	0.43 **	-0.42 **	-0.33 ***	-0.53 *
Mental health	-0.19	0.23	0.18	-0.15	-0.27	-0.11
Vitality	-0.41 **	0.36 ***	0.38 **	-0.39 **	-0.45 **	-0.48 **
Bodily pain	-0.28	0.25	0.34 **	-0.25	-0.18	-0.25
General health perception	-0.47 **	0.44 **	0.41 **	-0.38 **	-0.35	-0.39 **
Health change	-0.21	0.26	0.11	-0.24	-0.15	0.32 ***

* $P = <0.001$; ** $p = 0.001 - 0.008$; *** $p = 0.01 - 0.02$

Table 5.10 The Correlation between the five scales assessed in the study (p = <0.001)

	SNRS	FIM	CAMBS - Disability	CAMBS - Handicap	AI
EDSS	-0.92	-0.87	0.82	0.62	0.68
SNRS	-	0.87	-0.80	-0.58	-0.67
FIM	-	-	-0.85	-0.65	-0.73
CAMBS – Disability	-	-	-	0.65	0.54
CAMBS - Handicap	-	-	-	-	0.55

Table 5.11 Correlation between the EDSS, SNRS, Ambulation Index, and CAMBS and the patients' ability to work, do their housework, look after themselves and their perceived disability rank

	EDSS	SNRS	FIM	AI	CAMBS - Disability	CAMBS - Handicap
Work	0.69 *	-0.69 *	-0.59 *	0.59 *	0.66 *	0.48 *
House work	0.57 *	-0.59 *	-0.64 *	0.55 *	0.59 *	0.62 *
Independence	0.34 ***	-0.43 **	-0.44 **	0.35 ***	0.32 ***	0.36 ***
Disability rank	0.89 *	-0.92 *	-0.83 *	0.88 *	0.87 *	0.67 *

* $P = <0.001$; ** $p = 0.001 – 0.002$; *** $p = 0.01–0.02$

Slightly weaker but statistically significant correlations were also found between the SF-36 physical role limitation item and the EDSS (r = -0.50), SNRS (r = 0.46), FIM (r = 0.36), AI (r = -0.52), and the handicap domain of the CAMBS (r = -0.54); the SF-36 general health perception item and the EDSS (r = -0.47), SNRS (r = 0.44), the FIM (r = 0.41), AI (r = -0.38), and the handicap domain of the CAMBS (r = 0.39); the SF-36 social functioning item and the EDSS (r = -0.47), SNRS (r = 0.37), FIM (r = 0.43), AI (r = -0.42), and the disability and the handicap domains of the CAMBS (r = -0.33 and –0.53 respectively); the SF-36 vitality item and the EDSS (r = -0.41), SNRS (r = 0.36), FIM (r = 0.38), AI (r = -0.39), and the disability and the handicap domains of the CAMBS (r = -0.45 and -0.48 respectively), and between SF-36 bodily pain item and the FIM (r = 0.34) (Table 5.9). The correlation between the five scales assessed in the study is reported in Table 5.10.

B). Group differences and hypothesis testing

The two disability rank lists, which were compiled by myself and the research nurse, were almost identical (r = 0.99). All five scales, particularly the SNRS, correlated highly with the mean ranks of disability (Table 5.11). Each of the scales also correlated moderately with the patients' ability to work and do their house work, and weakly with the degree of patient's independence.

5.4 Discussion

Multiple sclerosis is a multifaceted disease characterised by a wide variability of clinical manifestations and natural history. Clinical rating scales used in this illness require relevant scale items, need to be able to embrace the whole range of affected domains, and should have high levels of reliability, validity, and responsiveness. Face and content validity of the currently existing scales have already been addressed by many researchers (Sharrack and Hughes, 1996) and reviewed by me in chapter 3. This study was designed to assess comprehensively the other psychometric properties of these scales.

5.4.1 EDSS

As reported by other researchers (Willoughby and Paty, 1988; Goodkin et al., 1989; Koziol et al., 1996), I found the frequency distribution of the EDSS scores to be bimodal with relative paucity of the middle scores. This bimodality

is unlikely to have been artefactual, despite the relatively small number of patients in this study, given the concordance between my findings and those obtained in cross-sectional studies of large population-based incident cohorts (Rodriguez et al., 1994; Midgard et al., 1996). Inter- and intra-rater reliabilities of the Functional System scores were comparably high. Similar to the previously reported studies, a difference of 2 points on the various Functional System scales achieved 97–100% rater agreement (Amato et al., 1988; Noseworthy et al., 1990; Goodkin et al., 1992) . Inter- and intra-rater reliabilities of the EDSS scores were equally high. Complete intra-rater and 96% inter-rater agreements were obtained by allowing a difference of 1.0 point (two 0.5 EDSS steps). These inter-rater reliability results are generally in accordance with the previously reported studies (Amato et al., 1988; Noseworthy et al., 1990; Goodkin et al., 1992), although Francis and co-workers reported the score difference between the two raters to vary between 2.0–4.0 points (four to eight 0.5 EDSS steps) in 10% of cases (Francis et al., 1991). Compared with my results, Goodkin and co-workers (1992) reported very high intra-rater reliability in a group of 10 patients with EDSS scores of 1-3.5, with complete intra-rater agreement obtained by allowing a difference of 0.5 points (one EDSS step). The discrepancy between these results and mine is likely to be due to the differences in the level of disease severity between the two cohorts, and to the time between the first and the second assessments. The assessments in my study were separated by three months whereas in the Goodkin study they were done on the same day and a practice effect cannot therefore be totally excluded. Although part of the difference between the first and the second assessments' scores in my study might have been due to a real but unreported change in the patients' clinical status, patients' variability between the two assessments was greatly minimised by including only those patients in whom clinical 'stability' was reported by both the patient and the rater.

The EDSS and its associated Functional Systems, with the exception of the mental Functional System, were insensitive to clinical change. The responsiveness of the EDSS has not been assessed previously in a manner exactly comparable to that used in my study. However, Ellison and co-workers (1993) found the Disability Status Scale (the previous version of the EDSS) to be insensitive to worsening of patient's clinical status as judged by the treating

neurologist, and Hobart and co-workers (1996d) found the EDSS to be unresponsive in a group of 64 patients with moderate to severe disability (EDSS 5.0-9.0). The face validity of the EDSS as a measure of combined impairment and disability was confirmed by its high correlation with the SNRS (particularly at the lower EDSS grades), the FIM, patients' disability ranks, and patients' self-assessment of disability using the physical functioning domain of the SF-36, and its moderate correlation with the Barthel Index. Similar high correlation with the physical functioning domain of the SF-36 has recently been reported by Rothwell and co-workers (1997), but sub-group analysis of the patients with EDSS scores of 5.0-7.5 in my cohort failed to replicate the high correlation between the EDSS and the Barthel Index reported by Hobart and co-workers (1996c) in 66 patients with moderate to severe disability (EDSS 5.0-9.0). As expected for any impairment / disability scale, the EDSS correlated moderately with measures of handicap and quality of life, and with patients' ability to work and do their housework.

5.4.2 SNRS

Contrary to the report of Koziol and co-workers (1996) that the SNRS has a near normal frequency distribution, I found the SNRS scores to be skewed to the 'normal' end of the scale with an additional cluster at the 'severely impaired' end of the scale suggesting both 'floor' and 'ceiling' effects. This discrepancy might, at least partially, be due to the differences in the range of disease severity, as assessed by the EDSS, between the two studies (2.0–8.0 in the Koziol study and 0–9.5 in my study). The internal consistency of the scale items was found to be surprisingly very high given the multidimensional nature of the scale suggesting a degree of item redundancy (Streiner and Norman, 1995c). Factor analysis showed the scale items, with the exception of the lower limb sensory and cerebellar items, to have segregated into five relatively 'meaningful' factors which explained most of the variance. Inter- and intra-rater reliabilities of the different SNRS items were variable, ranging between poor to substantial, depending on the definition of agreement. Although complete inter- and intra-rater agreements were only obtained by allowing a difference of 19 and 14 points respectively, partial agreement corrected reliability of the sum scores was high. A difference of 10 points only achieved 76% intra-rater and 85% inter-rater agreement, but a

difference of 13 points achieved more than 95% inter- and intra-rater agreement. These reliability results are similar to those published from the Scripps clinic (Koziol et al., 1996).

Despite this high reliability, neither the SNRS sum score nor any of its items were sensitive to clinical change. The only exception was the mentation and mood item, which was responsive mainly on the account of mood changes. The responsiveness of the SNRS was assessed in one previous study (Koziol et al., 1996), in which score changes on this scale were found to be 'more gradual' in comparison to those on the EDSS. Although direct comparison between this study and mine is not possible because of the methodological differences, I found both these scales to be unresponsive. The face validity of the SNRS as an impairment measure was supported by its high correlation with the EDSS. Surprisingly, the SNRS correlated highly with patients' disability ranks, the FIM, the disability domain of the CAMBS, and the physical functioning domain of the SF-36, but, as expected for an impairment scale, only moderately with the Barthel Index, patients' ability to work and do their house work, and other measures of handicap and quality of life.

5.4.3 FIM

The FIM sum scores were severely skewed to the 'less disabled' end of the scale, with a smaller cluster at the 'severely disabled' end of the scale suggesting both 'ceiling' and 'floor' effects. In a study of 201 patients with moderate to severe disability (EDSS 5.0 to 9.0), van der Putten and co-workers (1999) found the FIM sum scores to have span the entire scale range and to have a near Normal distribution with small 'ceiling' and 'floor' effects. The discrepancy between these results and mine is likely to be related to the differences in the level of disease severity between the two cohorts. As reported by Brosseau and Wolfson (1994), I found the internal consistency of the scale items to be surprisingly very high given the multidimensional nature of this scale suggesting a degree of item redundancy (Streiner and Norman, 1995c). Factor analysis suggested two factors which segregated the 'mobility' and the 'cognitive' items of the scale and accounted for most of the variance (although the latter accounted for only 6.4% of the total variance). Inter- and intra-rater reliabilities of the mobility items were generally comparably high. In comparison, the cognitive items were more reliable when

applied by the same rater reflecting their ambiguity and the lack of precision in differentiating between their different grades. Although complete inter- and intra-rater agreements were only obtained by allowing a difference of 13 and 9 points respectively, partial agreement corrected reliability of the FIM sum scores was high, and a difference of 9 points achieved more than 95% inter-rater agreement. These findings are consistent with other published reliability results in which the inter-rater reliability for the sum scores was found to be comparably high, and the inter-rater reliability of the mobility items to be higher than that of the cognitive items (Hall et al., 1993; Brosseau and Wolfson, 1994). There are no published data addressing the intra-rater reliability of this scale. As reported by others (Hall et al., 1993; van der Putten et al., 1999), I found the responsiveness of the FIM sum score and many of its mobility items to be high, thereby supporting the usefulness of this scale in clinical trials of MS. The face validity of the FIM as a disability scale was supported by the high correlation between this scale and other disability scales particularly the EDSS, the Barthel index (which is not surprising given the generic similarities between the two scales), the disability domain of the CAMBS, patients' self-assessment of disability using the physical functioning domain of the SF-36, and patients' disability ranks, and its moderate correlation with the AI. As expected for any disability scale, the FIM correlated moderately with other handicap and quality of life scales, but surprisingly highly with the SNRS.

The usefulness of the FIM in clinical trials of MS should be considered in the light of its limited content validity (Sharrack and Hughes, 1996). This scale is not comprehensive to the potential disabilities which could occur in this illness as it does not rate visual, speech, swallowing, affective, or sexual disabilities. It is my impression that the cognitive items of the FIM are the least useful part of this scale in the context of multiple sclerosis as they have a poor inter-rater reliability, are unresponsive, and explain only 6.4% of the total variance (factor analysis: mental factor). The FIM is also a somewhat cumbersome scale which requires reference to a 48-page instruction book and training for its application.

5.4.4 AI

As reported by other workers (Goodkin et al., 1989; Swingler and Compston, 1992), the frequency distribution of the AI scores was found to be bimodal with relative paucity of scores 7 and 8.

Rater reliability of this scale was very high. Complete inter-rater and 94.3% intra-rater agreement obtained by allowing a difference of a single point. These results are similar to the previously reported inter-rater reliability in a group of 20 patients which suggested that 95% of the raters would score within 1 point of the 'correct' score (Francis et al., 1991). No other rater reliability studies on this scale have been reported in the literature. Despite its high reliability, the AI was found to be weakly sensitive to clinical change. This is not surprising since this scale addresses only one dimension of the potential disabilities which could occur in this illness. The face validity of the AI as a disability scale was supported by its high correlation with patients' disability ranks, and patients' self-assessment of disability using the physical functioning domain of the SF 36. The AI correlated moderately with the EDSS, the FIM, and the Barthel index, reflecting its mono-dimensional nature in relation to the other disability scales. As expected for a disability scale, the AI correlated moderately with impairment, handicap, and quality of life scales.

5.4.5 CAMBS

As reported in its original publication (Mumford and Compston, 1993), the frequency distribution of the relapse and progression domains of this scale was skewed to the 'normal' end of the scales suggesting a 'floor' effect, which reflected the natural history of the illness and the patient population used in the study. In comparison, the frequency distribution of the disability domain was skewed to the 'severely disabled' end of the scale suggesting a 'ceiling' effect, whereas the handicap domain was evenly distributed. Rater reliability of this scale's four domains was reasonably high. Complete inter- and intra-rater agreement was obtained by allowing a difference of 1-2 points on the various domains. With the exception of handicap domain, a difference of 1 point achieved more than 95% rater agreement. The only published reliability data on this scale are those of Mumford and Compston (1993) who suggested a moderate reliability of the combined scores (calculated as kappa coefficient < 0.66) which is somewhat higher than what I found (kappa coefficient of 0.41). Nevertheless, these figures are of doubtful significance since a sum score is not used in the scale. The scale's relapse and progression domains were moderately sensitive to clinical change, reflecting their simple definition and the design of my

responsiveness study. The disability domain was weakly sensitive to clinical change, whereas the handicap domain was unresponsive.

The face validity of the disability domain of the CAMBS was supported by its high correlation with patients' disability ranks, the FIM, the EDSS, and the patients' self-assessment of disability using the physical functioning domain of the SF-36, and its moderate correlation with the Barthel index. This domain also correlated moderately with other impairment, handicap, and quality of life scales. Surprisingly, the correlation between the handicap domain of the CAMBS and the London Handicap Scale, patients' independence, and patients' ability to work and do their housework, was only moderate or weak, whereas its correlation with the EuroQol VAS was high thereby throwing doubt on the validity of this domain as a handicap scale. The previously reported high correlation between this domain and the Nottingham Health Profile was based on a study of only 10 patients (Mumford and Compston, 1993).

5.4.6 Raters' and patients' bias

This study was designed to minimise the effect of raters' and patients' bias on the assessment of reliability and responsiveness. All the raters were blinded to their own and other raters' previous scores, and open discussions about patients' clinical conditions were avoided amongst themselves. In the inter-rater reliability study, patients were assessed independently by the three raters and no fixed order for the examination was observed. The latter was designed to reduce the effect of patients' bias due to fatigue or recall of answers to specific questions required to obtaining some of the clinical scores. The effect of these potential sources of bias is unlikely to have been significant since the inter-rater reliability figures were not consistently higher than the intra-rater reliability figures which were based on assessments separated by three months periods. For the same reason, it is also unlikely that the familiarity of the patients with the assessors have increased the intra-rater reliability on the account of raters' recall of their previous scores, since such figures were often lower than the inter-rater reliability figures which were obtained in the same day by comparing the scores of two different raters.

Although inter-rater reliability was assessed on scores obtained by either two neurologists (EDSS, SNRS and AI) or a neurologist and a research nurse (FIM and CAMBS), it is unlikely that this has affected the validity of cross-scale

reliability comparison since the application of the FIM and the CAMBS was based on patients interview rather than neurological examination, and since previous studies have found these two scales to be equally reliable when applied by a neurologist and a nurse (Mumford and Compston, 1993), two therapists (Brosseau and Wolfson, 1994), or a neurologist and a multidisciplinary team comprising a doctor, an occupational therapist, a physiotherapist, a speech therapist, and a nurse (Kidd et al., 1995). Furthermore, the reliability of these scales was comparable when they were applied by the same neurologist twice (in the intra-rater reliability study), or by a neurologist and a research nurse (in the inter-rater reliability study).

A degree of rater bias was observed in the application of the SNRS reflecting the lack of clear guidelines for assessing the severity of impairment in this scale (Sharrack and Hughes, 1996). The reliability confidence intervals, particularly intra-rater, were relatively wide reflecting the small number of patients recruited for this study, as standard errors of measurements, used to construct the 95% confidence intervals, are inversely related to the sample size (Norman and Streiner, 1993b).

5.5 Conclusions

This study has comprehensively assessed the psychometric properties of five commonly used clinical scales in MS research. None of these scales completely satisfied the requirements of an ideal outcome measure, although many were found to have some desirable properties. The EDSS was reliable within 1.0 point (two 0.5 steps), valid as an impairment and disability scale, but not responsive. The SNRS was internally consistent, reliable within 13 point, valid as an impairment scale, but not responsive. The FIM was internally consistent, reliable within 9 point, valid as a disability scale, sensitive to clinical change, but had a limited content validity. The AI was reliable within 1 point, valid as an ambulation-related disability scale, but weakly sensitive to clinical change. The CAMBS was generally reliable within 1 point in each of its domains, and had valid disability and responsive relapse and progression domains. These results should inform the choice of outcome measures in future multiple sclerosis treatment trials.

Chapter 6

THE GUY'S NEUROLOGICAL DISABILITY SCALE

6.1 Introduction

The recent therapeutic trials in multiple sclerosis have highlighted the importance of measuring clinical outcomes but at the same time illustrated the inadequacies of the measures currently available. The need for a new outcome measure has become even more imperative with the increasing number of partially effective therapeutic agents which need to be assessed at a very high cost (Thompson and Noseworthy, 1996). In 1994, the U.S. National Multiple Sclerosis Society sponsored an international workshop entitled: "Outcome measures in multiple sclerosis clinical trials: a critical analysis" (Whitaker et al., 1995). This workshop was attended by over 100 investigators, statisticians, and other health care professionals who concluded that none of the current scales was adequate and that there was a need for the development of a new scoring system. Similar consensus was reached in other international workshops including the Second Algero Workshop on Multiple Sclerosis which was held in 1995 (Sharrack and Hughes, 1999b).

6.2 The need for a new outcome measure

To affirm the need for a new clinical scale for multiple sclerosis and to investigate its desirable properties, I conducted a postal survey among 49 leading neurologists from Europe, North America, Australia and New Zealand. The names of these neurologists were retrieved from the referee database of the Journal of Neurology, Neurosurgery and Psychiatry according to their clinical or research interest in multiple sclerosis and health measurement scales. All the participants were sent a questionnaire designed to ascertain whether they thought that existing clinical outcome measures in multiple sclerosis were adequate and whether there was a need for a new scale. They were also asked to indicate whether such a scale

should be patient or doctor orientated, ordinal or quantitative, uni- or multidimensional, and whether it should be biased towards any particular disability.

Thirty-five participants (71.4%) completed and returned their questionnaires. The majority of the respondents felt that currently existing outcome measures in multiple sclerosis were inadequate (85%) and that there was a need for a new scale (97%) which should be multidimensional (83.3%), ordinal (40%), patient orientated (46.6%), and not biased towards any particular disability (56.7%) (Table 6.1).

Table 6.1 The need for a new outcome measure for multiple sclerosis: results of the postal survey ($n = 35$)

Questions	**Results**			**Missing data ***
1. Do you think that currently available clinical outcome measures for multiple sclerosis are adequate?	*Yes* 13.3%	*No* 83.4%		3.3%
2. Is there a need for a new outcome measure for multiple sclerosis?	*Yes* 96.7%	*No* 0%		3.3%
3. Should the new outcome measure be patient or doctor orientated?	*Patient orientated* 46.6%	*Doctor orientated* 23.3%	*Both* 20%	10%
4. Should the new outcome measure be ordinal or quantitative?	*Ordinal* 40%	*Quantitative* 30%	*Both* 13.3%	16.7%
5. Should the new outcome measure be uni- or multi-dimensional?	*Uni-dimensional* 0%	*Multi-dimensional* 83.3%		16.7%
6. Should the new outcome measure be biased towards any particular disability?	*Yes* 20% **	*No* 56.7%		23.3%

* Some of the respondents left part of the questionnaire unanswered; ** Mainly towards lower limb disability

In response to these needs and to the growing consensus that disability should be the main focus of such a measure (Thompson and Hobart, 1996b), I undertook to develop a novel clinical disability scale for multiple sclerosis, which I have called the Guy's Neurological Disability Scale (GNDS).

6.3 GNDS conceptual model and development

Of the three dimensions of human disablement (impairment, disability, and handicap), disability is the main disease consequence that has a direct and practical relevance to patients as it determines their ability to perform their various daily activities. It is equally important to health services as it defines the level of care needed by the affected individuals and to society at large as it defines the social and economic impact of the illness on the affected individuals. Disability is therefore the common pathway by which the diverse consequences of complex diseases such as multiple sclerosis can be assessed.

The GNDS is based on the concept that disability in multiple sclerosis is multidimensional, and can be considered in several separate categories (Sharrack and Hughes, 1999a). A list of the various disability domains relevant to patients with multiple sclerosis was compiled from a review of previously published morbidity literature (Swingler and Compston, 1992; Rodriguez et al., 1994; Midgard et al., 1996) and existing outcome measures (Sharrack and Hughes, 1996), supplemented by open interviews with 5 patients with multiple sclerosis. Twelve separate categories of mutually exclusive human functions capable of capturing all the aspects of disability which could be encountered in multiple sclerosis were identified. These categories included cognition, mood, vision, speech, swallowing, upper limb function, lower limb function, bladder function, bowel function, fatigue, sexual function, and 'others' to cover disabilities resulting from dysfunction of other less defined systems such as pain, spasms, vertigo, etc. These 12 dimensions were later confirmed independently by Thompson and Hobart (1996b) as being the most commonly affected aspects of human function in patients with multiple sclerosis.

This model differed conceptually from the EDSS Functional Systems in that it divided some of these systems into their constituent parts (sphincters into bladder, bowel, and sexual function; cerebral into cognition and mood; and brain stem into speech and swallowing), incorporated others into regional functions (pyramidal, cerebellar and sensory are replaced by upper and lower limb function), kept others as separate categories (vision, others), and created a new category for fatigue which is often a troublesome and a disabling problem in multiple sclerosis (Table 6.2).

Table 6.2 Conceptual correlation between the GNDS categories and EDSS Functional Systems

EDSS Functional Systems	GNDS dimensions
Cerebral	Cognition, Mood
Visual	Visual
Brain stem	Speech and communication, Swallowing
Pyramidal, Cerebellar, Sensory	Upper limb, Lower limb
Sphincter	Bladder, Bowel, Sexual
-	Fatigue
Other	Other

The new model also differed conceptually from other scales of activities of daily living, such as the Barthel Index and the FIM, in that such activities were considered at their basic anatomical levels to avoid any bias towards upper and lower functions (Table 6.3). It also had the advantage of being able to cover the whole range of activities of daily living without the need to name them individually or risk selection bias.

The level of disability in each one of these 12 categories was graded according to its severity and its impact on patients as judged by the help required to perform these functions according to a 7-level scoring system which was adapted from the WHO ICIDH disability severity scale (World Health Organisation, 1980). This scoring system ranged from 0 (normal status) to 6 (total loss of function) (Table 6.4).

The range of the possible disabilities which could be encountered in each one of the 12 dimensions and the stages at which they occur in relation to the disease severity and duration were based on detailed review of the literature supplemented by input from 10 experts in neurology, neuropsychology, neurourology, and health status measurements at the Guy's, King's and St. Thomas' School of Medicine (Professor RAC Hughes, Professor J Weinman), the National Hospital for Neurology and Neurosurgery (Professor A Thompson, Professor M Ron, Dr. G Plant, Dr. C Fowler), the Institute of Psychiatry (Professor G Dunn, Dr. S Wessely), the Rivermead Rehabilitation Centre (Dr. D Wade), and the University of Wales College of Medicine (Professor M Wiles).

Table 6.3 Conceptual correlation between the GNDS categories and other scales of activities of daily living

Barthel Index	FIM	GNDS
Bladder, Bowel	Bladder, Bowel	Bladder, Bowel
Grooming	Grooming	Upper limb function
Toilet use	Self-care: toileting, Mobility: transfer to toilet	Upper limb function, Lower limb function,
Feeding	Eating	Upper limb function, Swallowing
Transfer	Transfer: bed/chair, tub/shower	Upper limb function, Lower limb function,
Mobility	Mobility: walking	Lower limb function
Dressing	Self-care: dressing-upper body, dressing-lower body	Upper limb function
Stairs	Mobility: stairs	Lower limb function
Bathing	Bathing	Upper limb function
-	Communication: comprehension, expression	Speech, Cognition
	Social cognition: social interaction, problem solving, memory	Cognition, Mood
-	-	Vision
-	-	Swallowing
-	-	Sexual function
-	-	Fatigue
-	-	Others

Table 6.4 GNDS severity grades

Grade	Level of disability
0	Normal function
1	Symptoms of no consequences
2	Symptoms causing difficulties, but not disabilities
3	Disability, but no help is required
4	Disability requiring help
5	Almost total loss of function
6	Total loss of function
X	Unknown

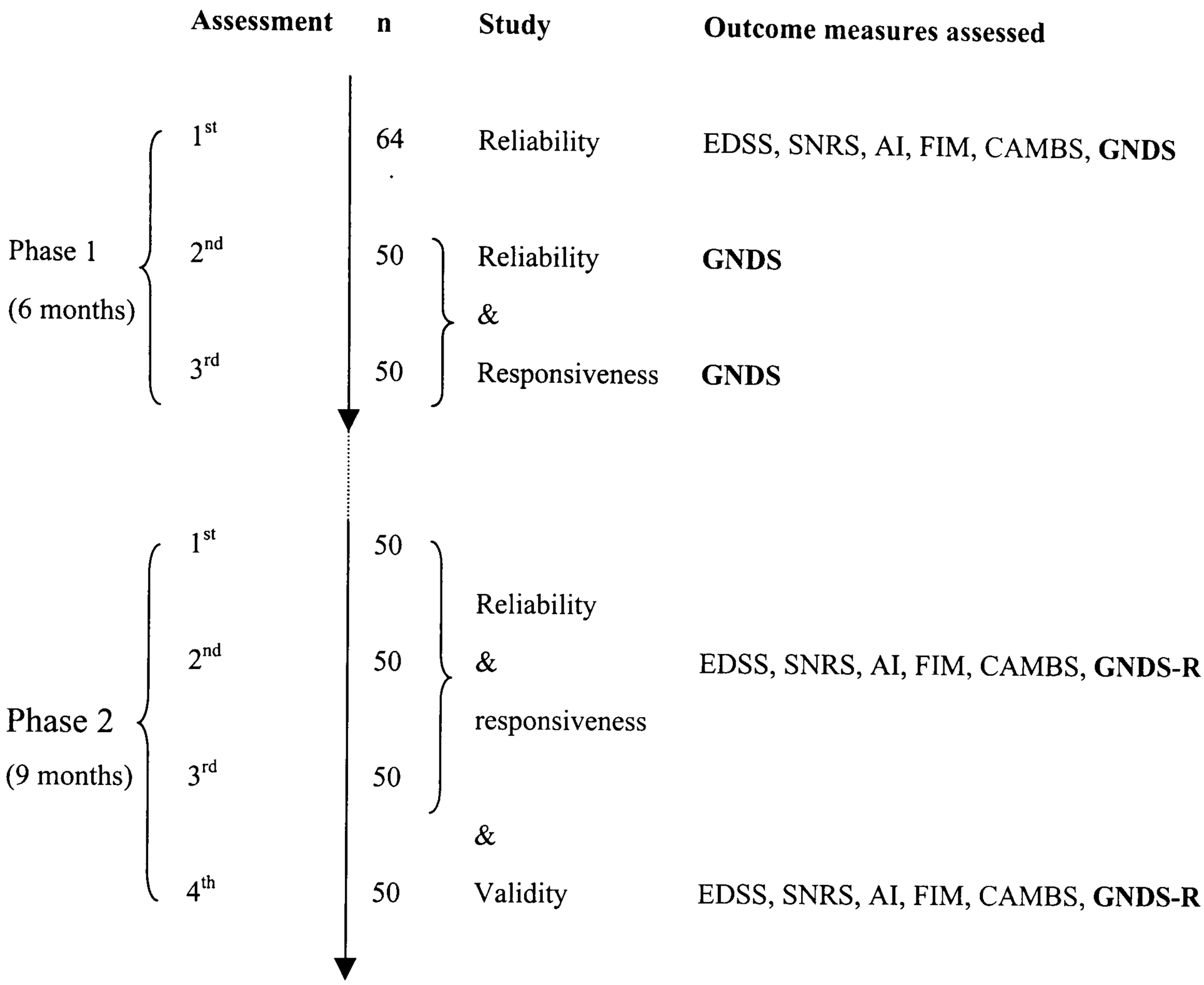
This information was utilised to create 12 separate disability sub-scales in which the range of individual disabilities was graded according to their severity. These grades were arranged so that each step represented as far as possible the same level of disability in each dimension according to the GNDS severity scoring system. To optimise the reproducibility of the GNDS, each sub-scale was further supplemented by an interview section which consisted of a set of questions designed to ascertain the presence and the severity of individual disabilities as outlined in the relevant disability sub-scale. Individual disability scores could thereafter be deduced from the patient's answers according to the relevant 'scoring sections' and an overall score, describing the patient's total disability, could be reached by summing all the different sub-scores giving a sum score ranging between 0 (no disability) and 72 (maximum possible disability). This process generated a 12-item scale.

The face and content validity of GNDS were assessed during the international postal survey described earlier by inviting the same cohort of 49 neurologist to review the scale and indicate the degree of their approval or disapproval using a standard questionnaire. The referees were also invited to provide constructive criticisms and suggestions to improve the scale and its different items. The results of this study showed that the majority of the referees (73%) approved the scale therefore confirming its face validity. However many referees provided critical comments in relation to the scale contents which suggesting that the scale needed to be revised.

The GNDS was piloted on the same cohort of 64 patients who took part in the inter-rater reliability study of the five commonly used multiple sclerosis scales (described in chapter 5) to assess its inter-rater reliability and construct validity, and on the sub-group of 50 patients who took part in the intra-rater reliability and responsiveness study (described in chapter 5) who were followed up for 6 months to assess its intra-rater reliability and responsiveness (Figure 6.1). The study design was identical to the one described in chapter 5. This pilot study showed the GNDS to be internally consistent (Cronbach's alpha of 0.79), have high inter- and intra-rater reliability (intraclass correlation coefficients 0.99 and 0.96 receptively), and to be moderately responsive to clinical change (effect size 0.54, $p = <0.001$). However the frequency distribution of the GNDS scores was skewed

to the 'less disabled' end of the scale suggesting a floor effect. The GNDS therefore needed to be revised.

Figure 6.1 GNDS / GNDS-R study design



n = number of patients in the study

6.4 The Revised GNDS

The critical comments of the 35 referees who responded to the GNDS face and content validity questionnaire were categorised into general comments related to the conceptual model of the scale, its general structure, and its dimensions, and specific comments related to its 12 sub-scales. Taking into account the general comments, the GNDS severity scoring system was modified by reducing the disability steps from 7 to 6 (Table 6.5).

Table 6.5 GNDS-R severity grades

Grade	Level of disability
0	Normal status
1	Symptoms causing no disability
2	Mild disability - not requiring help from others
3	Moderate disability - requiring help from others
4	Severe disability - almost total loss of function
5	Total loss of function - maximum help required

The specific comments were utilised to revise the GNDS creating the Revised Guy's Neurological Disability Scale (GNDS-R). These revisions resulted in a new scale which specified 60 points between 0 (no disability) and 60 (maximum possible disability) (Appendix 1).

6.5 The psychometric evaluation of the GNDS-R

The changes suggested by the referees were of such a magnitude as to have resulted in a scale which is substantially different from the original version. The psychometric properties of the revised scale therefore needed to be re-assessed again, ideally on a naive cohort of patients to avoid any bias resulting from training effects (Peto et al. 1995). However due to administrative difficulties in recruiting and following a second cohort of patients, the evaluation was conducted on the same cohort of 50 patients who took part in the intra-rater and responsiveness assessment of the GNDS and the five commonly used multiple sclerosis scales (described in chapter 5) (Table 6.1). This strategy was felt to be appropriate since the revisions have effectively resulted in a new scale, and since the frequent administrations of the GNDS during the intra-rater and responsiveness study were separated by three monthly intervals which minimised any subject or rater bias resulting from the effect of training or recall of previous answers.

6.5.1 Face validity

Copies of the revised scale were posted to the same group of 49 international referees who took part in the first face and content study. The

referees were invited to critically study the GNDS-R and indicate the degree of their approval or disapproval of the scale and its 12 categories using a 6-point scoring system (Table 6.6).

Thirty-three referees (67.3%) responded to the questionnaire, 4 of whom left parts of the questionnaire unanswered. As compared with the first face validity study, more referees strongly approved / approved of the revised scale. Eighty two percent of the respondents expressed their approval of the scale in general. The majority of the responders also approved the cognitive (88%), mood (73%), vision (85%), speech (85%), swallowing (88%), upper limb (85%), lower limb (85%), bladder (91%), bowel (94%), fatigue (82%), sexual function (82%) and the 'others' (79%) sub-scales (Table 6.6). These results support the validity of the GNDS-R as a disability scale for multiple sclerosis.

6.5.2 Inter-rater reliability

Inter-rater reliability was assessed on the same cohort of 50 patients with multiple sclerosis described in the in intra-rater reliability and responsiveness study in chapter 5. This cohort consisted of 31 women and 19 men with a median age of 36 years (range 24–51), median EDSS score of 4.5 (range 0-7.5), and a median disease duration of 12 years (range 2-17). All patients were assessed independently at the same session by myself and a second neurologist who was familiar with the GNDS-R. To prevent any systematic bias resulting from practice effect or fatigue, no fixed order for the examination of the patients by each rater was observed. The median (range) of the GNDS-R sum and sub-scores of the two assessments were almost identical (Table 6.7).

The frequency distribution of the sum scores of both raters was slightly skewed to the normal end of the scale (Figure 6.2), which reflects the mild to moderate degree of disability in this cohort.

Table 6.6 Face validity of the GNDS-R

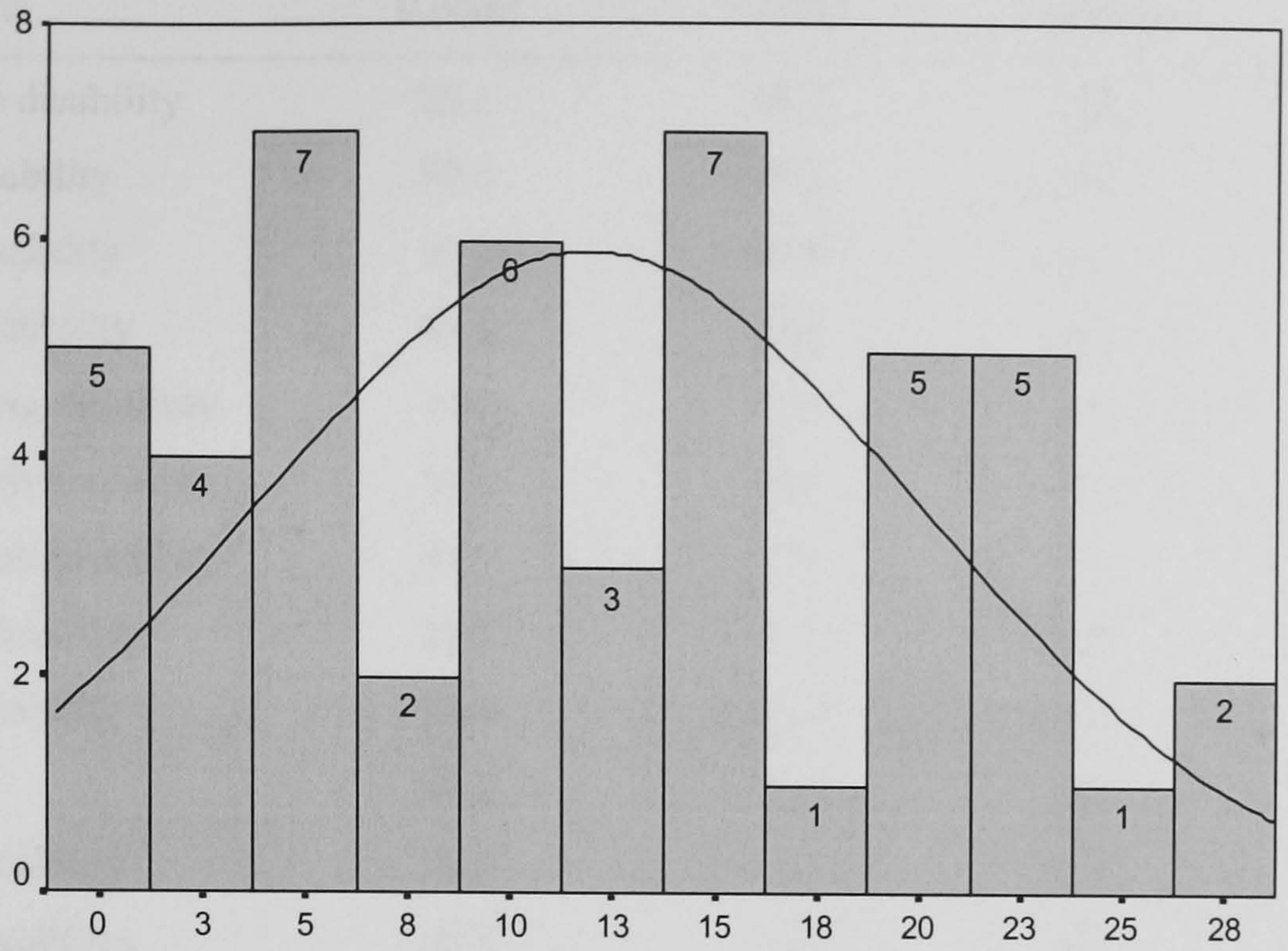
Scale item	% of respondents (n = 33)						
	Strongly approve	Approve	Tends to approve	Tends to disapprove	Disapprove	Strongly disapprove	Missing data *
Cognition	15.2	48.5	24.2	6.1	3	3	0
Mood	9.1	54.5	9.1	21.1	0	6.1	0
Vision	30.3	33.3	21.1	6.1	6.1	3	0
Speech	24.2	42.4	18.2	9.1	6.1	0	0
Swallowing	30.3	48.5	9.1	9.1	3	0	0
Upper limb	21.1	39.4	24.2	12.1	3	0	0
Lower limb	24.2	39.4	21.1	3	3	0	0
Bladder	24.2	48.5	21.1	3	0	6.1	0
Bowel	24.2	48.5	21.1	3	3	0	0
Fatigue	24.2	48.5	9.1	15.2	3	0	0
Sex function	21.1	39.4	21.1	12.1	6.1	0	0
Others	12.1	45.5	21.2	3	3	3	12.1
Overall scale	15.1	45.5	21.2	9.1	6.1	0	3

* Some of the referees left part of the questionnaire unanswered.

Table 6.7 Median (range) scores of the GNDS-R inter-rater reliability study (n = 50)

Scale item	Rater 1: Median (range)	Rater 2: Median (range)
Cognitive disability	0 [0 to 3]	0 [0 to 3]
Mood disability	0 [0 to 4]	0 [0 to 4]
Visual disability	0 [0 to 2]	0 [0 to 2]
Speech disability	0 [0 to 3]	0 [0 to 3]
Swallowing disability	0 [0 to 2]	0 [0 to 2]
Upper limb disability	1 [0 to 4]	1 [0 to 4]
Lower limb disability	2 [0 to 4]	2 [0 to 4]
Bladder disability	2 [0 to 4]	2 [0 to 4]
Bowel disability	0 [0 to 5]	0 [0 to 5]
Fatigue	2 [0 to 4]	2 [0 to 4]
Sexual disability	0 [0 to 5]	1 [0 to 5]
Other disabilities	1 [0 to 4]	1 [0 to 4]
GNDS-R sum score	12 [0 to 26]	12 [0 to 28]

Figure 6.2 The frequency distribution (with the superimposed normal curve) of the GNDS-R ($n = 50$).



The largest score difference between the two raters was 0 point for the swallowing disability sub-scale, 1 point for the speech, upper limb, and lower limb disability sub-scales, 2 points for the cognitive, mood, visual, bladder, bowel, fatigue, and other disabilities sub-scales, and 3 points for the sexual function disability sub-scale (Table 6.8).

Inter-rater agreement on the different GNDS-R sub-scales was high with kappa coefficients ranging between 0.54 and 1 (moderate to perfect), intraclass correlation coefficients ranging between 0.82 and 1 (almost perfect to perfect), and repeatability coefficients ranging between 0 and 1.5 points. Inter-rater agreement on the GNDS-R sum scores was 35.4%, 70.8%, 93.8%, and 100% when agreement was defined as no difference, a difference ≤ 1, 2, and 3 points respectively (Table 6.8), with a repeatability coefficient of 2.6 points, a kappa coefficient was 0.31 (moderate), and an intraclass correlation coefficient was 0.98 (almost perfect) (Table 6.9). The mean score differences between the two raters for all GNDS-R sub-scores and the sum score was very small with narrow 95% confidence intervals which included the '0' value suggesting the absence of any rater bias.

Table 6.8 GNDS-R inter-rater score agreement (%) (n = 50)

Scale items	0 point	1 point	2 points	3 points
Cognitive disability	89.6	95.8	100	NA
Mood disability	89.6	97.9	100	NA
Visual disability	89.6	97.9	100	NA
Speech disability	89.6	100	NA	NA
Swallowing disability	100	NA	NA	NA
Upper limb disability	89.6	100	NA	NA
Lower limb disability	91.7	100	NA	NA
Bladder disability	81.3	97.9	100	NA
Bowel disability	89.6	95.8	100	NA
Fatigue	81.3	97.9	100	NA
Sexual disability	79.2	91.7	97.9	100
Other disabilities	66.8	95.9	100	NA
GNDS-R sum score	35.4	70.8	93.8	100

Table 6.9 GNDS-R inter-rater reliability (n = 50)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Cognitive disability	0.79 [0.63 to 0.95]	0.90 [0.68 to 0.99]	0.15 [0.01 to 0.28]	0.9
Mood disability	0.82 [0.68 to 0.96]	0.93 [0.69 to 0.99]	0.13 [0.01 to 0.24]	0.8
Visual disability	0.82 [0.67 to 0.97]	0.85 [0.46 to 0.99]	0.04 [-0.16 to 0.08]	0.8
Speech disability	0.74 [0.52 to 94]	0.94 [0.79 to 0.99]	0.06 [-0.03 to 0.16]	0.6
Swallowing disability	1 N/A	1 N/A	0 N/A	0
Upper limb disability	0.86 [0.74 to 0.98]	0.96 [0.86 to 0.99]	-0.06 [-0.16 to 0.03]	0.6
Lower limb disability	0.89 [0.79 to 0.99]	0.98 [0.93 to 0.99]	0.04 [-0.04 to 0.13]	0.6
Bladder disability	0.74 [0.59 to 0.89]	0.92 [0.73 to 0.99]	-0.06 [-0.24 to 0.11]	1.1
Bowel disability	0.80 [0.66 to 0.94]	0.95 [0.80 to 0.99]	0.10 [-0.03 to 0.24]	0.9
Fatigue	0.74 [0.59 to 0.99]	0.94 [0.77 to 0.99]	0 [-0.15 to 0.15]	1
Sexual disability	0.70 [0.54 to 0.86]	0.89 [061 to 0.99]	0.06 [-0.28 to 0.16]	1.5
Other disabilities	0.54 [0.36 to 0.72]	0.82 [0.64 to 0.99]	0.10 [-0.09 to 0.29]	1.3
GNDS-R sum score	0.31 [0.16 to 0.46]	0.98 [0.95 to 0.99]	0.42 [0.04 to 0.79]	2.6

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

6.5.3 Internal consistency

Internal consistency was assessed using the inter-rater reliability data set. Cronbach's alpha of the GNDS-R was 0.79 indication satisfactory internal consistency. The deletion of individual sub-scales did not lead to a significant increase in the Cronbach's alpha suggesting that the various scale items were homogeneous (Table 6.10). Item-total correlation showed that none of the scale items correlated with the total score below 0.32. Split-half reliability showed the two randomly created halves of the scale to be highly correlated with each other with a Spearman-Brown correlation coefficient of 0.79.

Table 6.10 Internal consistency of GNDS-R

Scale item	Item-total correlation	Cronbach's alpha if item deleted
Cognitive disability	0.36	0.79
Mood disability	0.33	0.80
Visual disability	0.40	0.80
Speech disability	0.33	0.77
Swallowing disability	0.42	0.80
Upper limb disability	0.67	0.76
Lower limb disability	0.74	0.75
Bladder disability	0.56	0.77
Bowel disability	0.32	0.80
Fatigue	0.45	0.78
Sexual disability	0.38	0.80
Other disabilities	0.39	0.80

6.5.4 Intra-rater reliability and responsiveness

Intra-rater reliability and responsiveness of the GNDS-R were as assessed on same cohort of 50 patients with multiple sclerosis at the time of the intra-rater reliability and responsiveness study of the five commonly used multiple sclerosis scales which was described in chapter 5. The patients were followed for nine months with three monthly assessments. During each visit, patients underwent a full neurological examination, and were assigned scores on the SNRS, EDSS, FIM, AI, CAMBS, 10 metre walk, and the Barthel Index, and they were asked to complete the EuroQol health related quality of life visual analogue scale (VAS) (EuroQol Group, 1990). I also completed a separate copy of the EuroQol VAS to

reflect my subjective perception of the patient's health status. As with the intra-rater reliability and responsiveness study described in chapter 5, patients' overall status were classified as stable, improved, or worsened, if both the patients' and my assessments were identical indicating no change, improvement, or worsening respectively. Intra-rater reliability was tested on the pairs of assessment between which patients' overall status were judged to have remained stable, whereas responsiveness was tested on the pairs of assessment between which they had changed (improved or worsened).

A). Intra-rater reliability

Thirty-five patients had remained stable between two visits on at least one occasion during the 9 months follow up period. To avoid introducing any statistical bias, only one pair of assessments (the first) per patient was included in the final analysis. This cohort consisted of 20 women and 15 men with a median age of 38 years (range 24-51 years), a median EDSS of 4.5 (0-7.5), and median disease duration of 11 years (2-17 years). The median (range) of the GNDS-R sum and the sub-scores of the two assessments were identical (Table 6.11)

Table 6.11 Median (range) scores of the GNDS-R intra-rater reliability study ($n = 35$)

Scale item	Time 1: Median (range)	Time 2: Median (range)
Cognitive disability	0 [0 to 3]	0 [0 to 3]
Mood disability	0 [0 to 3]	0 [0 to 3]
Visual disability	0 [0 to 2]	0 [0 to 2]
Speech disability	0 [0 to 3]	0 [0 to 3]
Swallowing disability	0 [0 to 2]	0 [0 to 2]
Upper limb disability	1 [0 to 4]	1 [0 to 4]
Lower limb disability	2 [0 to 4]	2 [0 to 4]
Bladder disability	2 [0 to 4]	2 [0 to 3]
Bowel disability	0 [0 to 4]	0 [0 to 4]
Fatigue	2 [0 to 4]	1 [0 to 4]
Sexual disability	0 [0 to 4]	0 [0 to 4]
Other disabilities	0 [0 to 3]	0 [0 to 3]
GNDS-R sum score	9 [0 to 25]	9 [0 to 25]

The largest score differences between the two assessments were 1 point for the vision, lower limb, and sexual function sub-scales, 2 points for the cognition, speech, swallowing, and upper limb sub-scales, 3 points for the mood, bladder, fatigue, and the other disabilities sub-scales, and 4 points for the bowel sub-scale (Table 6.12).

Table 6.12 GNDS-R intra-rater score agreement (%) (n = 35)

Scale item	within 0 point	within 1 point	within 2 points	within 3 points	within 4 point	within 5 points
Cognitive disability	77.1	88.6	100	NA	NA	NA
Mood disability	57.1	85.7	94.3	100	NA	NA
Visual disability	74.3	100	NA	NA	NA	NA
Speech disability	91.4	97.2	100	NA	NA	NA
Swallowing disability	97.1	97.1	100	NA	NA	NA
Upper limb disability	74.3	97.2	100	NA	NA	NA
Lower limb disability	85.7	100	NA	NA	NA	NA
Bladder disability	80	94.3	94.3	100	NA	NA
Bowel disability	85.7	97.2	97.2	97.2	100	NA
Fatigue	68.6	85.8	97.3	100	NA	NA
Sexual disability	85.7	100	NA	NA	NA	NA
Other disabilities	71.4	85.7	97.2	100	NA	NA
GNDS-R sum score	22.9	60	85.7	91.5	91.5	100

Intra-rater agreement on the different sub-scales was variable, with kappa coefficients ranging between 0.46 and 0.87 (moderate to almost perfect), intraclass correlation coefficients ranging between 0.77 and 0.96 (substantial to almost perfect), and repeatability coefficients ranging between 0.7 and 1.8 points. Intra-rater agreement on the GNDS-R sum scores was 23%, 60%, 86%, 92%, 92%, and 100% when agreement was defined as no difference, a difference of ≤ 1, 2, 3, 4, and 5 points respectively, with a repeatability coefficient of 4.1 points, a kappa coefficient of 0.18 (poor), and an intraclass correlation coefficient of 0.96 (almost perfect) (Table 6.11 and 6.13). With exception of the bowel disability sub-scale, the mean score differences between the two assessments were small with relatively narrow 95% confidence intervals indicating the absence of any raters' bias.

Table 6.13 GNDS-R intra-rater reliability (n = 35)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Cognitive disability	0.64 [0.40 0.88]	0.80 [0.36 to 0.99]	0.06 [-0.31 to 0.21]	1.5
Mood disability	0.46 [0.24 to 0.68]	0.70 [0.31 to 0.99]	0.06 [-0.43 to 0.31]	2.1
Visual disability	0.61 [0.32 to 0.90]	0.91 [0.42 to 0.99]	-0.13 [0.21 to 0.15]	1
Speech disability	0.84 [0.56 to 0.99]	0.92 [0.26 to 0.99]	-0.03 [-0.21 to 0.15]	1
Swallowing disability	0.87 [75 to 0.99]	0.94 [0.59 to 0.99]	0.06 [-0.06 to 0.17]	0.7
Upper limb disability	0.61 [0.51 to 0.71]	0.87 [0.58 to 0.99]	-0.11 [-0.31 to 0.08]	1.1
Lower limb disability	0.80 [0.72 to 0.88]	0.96 [0.87 to 0.99]	0.09 [-0.04 to 0.21]	0.7
Bladder disability	0.76 [0.66 to 0.86]	0.89 [0.45 to 0.99]	0.14 [-0.13 to 0.42]	1.6
Bowel disability	0.51 [0.28 to 0.74]	0.81 [0.37 to 0.99]	0.71 [-0.09 to 0.43]	1.5
Fatigue	0.56 [0.36 to 0.76]	0.74 [0.26 to 0.99]	0.03 [-0.29 to 0.36]	1.9
Sexual disability	0.80 [0.61 to 0.99]	0.96 [0.91 to 0.99]	-0.14 [-0.28 to –0.02]	0.7
Other disabilities	0.59 [0.36 to 0.82]	0.77 [0.49 to 0.99]	0 [-0.33 to 0.33]	1.8
GNDS-R sum score	0.18 [0.01 to 0.35]	0.96 [0.93 to 0.99]	0.06 [-0.64 to 0.76]	4.1

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

B). Responsiveness

Of the 50 patients assessed, 25 were found to have changed on at least one occasion during the 9 months follow up period. This group consisted of 20 women and 5 men with a median age of 36 years (range 24-51), median EDSS of 5.5 (range 0–7.5), and a median disease duration of 10 years (range 2-22). To avoid introducing any statistical bias, only one pair of assessments (the first) per patient was included in the final analysis. The order of assessment in each pair (15 patients worsened, and 10 improved) was later re-arranged so as to make all the changes of one direction (stable or improved to worsened). Patients' subjective assessments of their own health status using the EuroQol VAS was moderately sensitive to clinical change with an effect size of 0.55, p <0.001. Similar subjective assessment by myself using the EuroQol VAS was weakly sensitive to clinical change with an effect size of 0.36, p <0.001. The GNDS-R sum score was sensitive to clinical change with an effect size of 0.58, p <0.001

(moderate). The mood, vision, upper limb, lower limb, bladder, fatigue, and other disabilities sub-scales were also sensitive to clinical change with effect size values ranging between 0.23, $p = 0.021$ (small) to 0.92, $p = 0.001$ (large). The other sub-scales (cognition, speech, swallowing, bowel, and sexual) were unresponsive in this cohort (Table 6.14).

Table 6.14 GNDS-R responsiveness ($n = 25$)

Scale item	Time 1 Median (range) score	Time 1 Median (range) score	p *	Effect size
Cognitive disability	0 [0 to 3]	0 [0 to 3]	0.185	0.19
Mood disability	0 [0 to 3]	1 [0 to 4]	0.023	0.55
Visual disability	0 [0 to 2]	1 [0 to 2]	0.011	0.49
Speech disability	0 [0 to 3]	0 [0 to 3]	0.863	0.04
Swallowing disability	0 [0 to 2]	0 [0 to 2]	1	0
Upper limb disability	1 [0 to 3]	1 [0 to 3]	0.014	0.30
Lower limb disability	1 [0 to 4]	2 [0 to 4]	0.021	0.23
Bladder disability	2 [0 to 4]	3 [0 to 5]	0.035	0.37
Bowel disability	0 [0 to 4]	0 [0 to 5]	0.035	0.15
Fatigue	0 [0 to 3]	2 [0 to 3]	0.002	0.92
Sexual disability	0 [0 to 4]	0 [0 to 5]	0.096	0.17
Other disabilities	0 [0 to 5]	1 [o to 4]	0.005	0.58
GNDS-R sum score	9 [0 to 23]	15 [1 to 30]	<0.001	0.58

* Wilcoxon Signed Ranks test

6.5.5 The reliability of the GNDS-R when applied by other health care personnels and by patients' close relatives or carers

To simplify and reduce the cost of clinical trials, the GNDS-R was designed as a simple scale which could be applied by any health care personnel. Inter-rater reliability of the scale when applied by a neurologist and a research nurse, or by a neurologist and a patient's relative or carer was therefore tested on the same cohort of 50 patients during the intra-rater reliability and responsiveness study. During the first of the three monthly visits, the GNDS-R was administered independently by myself and a research nurse in the same session. A patient's relative or close friend, unfamiliar with the scale, was also provided with a copy of the scale during the second of the three monthly visits, and was asked to

administer it at home within 72 hours and post it back to me. Records of the time needed to apply the scale by all the raters were kept to assess the burden of the scale application both on patients and raters.

A). Doctor-nurse administration

All 50 patients took part in this study. The median (range) of the GNDS-R sum and sub-scores of the neurologist's (mine) and the nurse's assessments were almost identical (Table 6.15).

Table 6.15 Median (range) scores of the GNDS-R doctor-nurse reliability study ($n = 50$)

Scale item	Doctor: median (range)	Nurse: median (range)
Cognitive disability	0 [0 to 4]	0 [0 to 4]
Mood disability	0 [0 to 3]	0 [0 to 3]
Visual disability	0 [0 to 2]	0 [0 to 2]
Speech disability	0 [0 to 3]	0 [0 to 3]
Swallowing disability	0 [0 to 2]	0 [0 to 2]
Upper limb disability	1 [0 to 4]	1 [0 to 4]
Lower limb disability	3 [0 to 5]	2 [0 to 5]
Bladder disability	0 [0 to 4]	0 [0 to 4]
Bowel disability	0 [0 to 5]	0 [0 to 5]
Fatigue disability	2 [0 to 5]	2 [0 to 5]
Sexual disability	0 [0 to 5]	0 [0 to 5]
Other disabilities	1 [0 to 3]	1 [0 to 3]
GNDS sum score	13 [0 to 28]	13 [0 to 28]

Inter-rater agreement on the different sub-scales was high, with kappa coefficients ranging between 0.58 and 0.95 (moderate to almost perfect), intraclass correlation coefficients ranging between 0.87 and 0.96 (almost perfect), and repeatability coefficients ranging between 0.2 and 1.7 point. Inter-rater agreement on the GNDS-R sum score was equally high with kappa coefficient of 0.25 (fair), an intraclass correlation coefficient of 0.96 (almost perfect), and a repeatability coefficient of 4.3 point (Table 6.16). The mean score differences between the two raters were small with relatively narrow 95% confidence intervals which included the '0' value indicating the absence of any raters' bias. The mean (SD) of the time required for scale administration and scoring was 4

minutes and 30 seconds (2 minutes) for the neurologist (myself) and 6 minutes and 23 seconds (4 minutes and 22 seconds) for the nurse.

Table 6.16 GNDS-R doctor-nurse inter-rater reliability (n = 50)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Cognitive disability	0.69 [0.49 to 0.89]	0.78 [0.34 to 0.99]	-0.20 [-0.46 to 0.05]	1.7
Mood disability	0.81 [0.66 to 0.96]	0.82 [0.45 to 0.99]	0.12 [0.04 to 0.28]	1.1
Visual disability	0.74 0.58 to 0.90]	0.87 [0.52 to 0.99]	0.06 [-0.05 to 0.17]	0.7
Speech disability	0.79 [0.62 to 0.96]	0.93 [0.76 to 0.99]	0 [-0.08 to 0.08]	0.6
Swallowing disability	0.93 [0.87 to 0.99]	0.97 [0.85 to 0.99]	-0.02 [-0.06 to 0.02]	0.3
Upper limb disability	0.95 [0.91 to 0.99]	0.98 [0.95 to 0.99]	-0.04 [-0.09 to 0.02]	0.4
Lower limb disability	0.89 [0.79 to 0.99]	0.98 [0.94 to 0.99]	-0.04 [-0.12 to 0.04]	0.6
Bladder disability	0.87 [0.76 to 0.98]	0.98 [0.95 to 0.99]	0 [-0.08 to 0.08]	0.2
Bowel disability	0.82 [0.67 to 0.96]	0.87 [0.57 to 0.99]	0.06 [0.11 to 0.26]	1.3
Fatigue	0.58 [0.42 to 0.74]	0.90 [0.67 to 0.99]	0.12 [-0.34 to 0.10]	1.5
Sexual disability	0.76 [0.58 to 0.94]	0.89 [0.62 to 0.99]	0.04 [-0.21 to 0.29]	1.7
Other disabilities	0.73 [0.58 to 0.88]	0.87 [0.60 to 0.99]	0 [-0.18 to 0.18]	1.2
GNDS sum score	0.25 [0.05 to 0.45]	0.96 [0.85 to 0.99]	0.16 [-0.79 to 0.47]	4.3

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

B). Doctor – relative / carer administration

Thirty-nine patients took part in this study. The other 11 patients did not have a close relative or a carer available to administer the scale in the specified time frame. The median (range) of the GNDS-R sum and sub-scores of the neurologist's (mine) and the patient relatives' assessments were very close (Table 6.17).

Table 6.17 Median (range) scores of the GNDS-R doctor-relative reliability study ($n = 39$)

Scale item	Doctor: median (range)	Relative: median (range)
Cognitive disability	0 [0 to 4]	0 [0 to 4]
Mood disability	0 [0 to 3]	0 [0 to 3]
Visual disability	0 [0 to 3]	0 [0 to 3]
Speech disability	0 [0 to 3]	0 [0 to 3]
Swallowing disability	0 [0 to 2]	0 [0 to 2]
Upper limb disability	1 [0 to 4]	1 [0 to 4]
Lower limb disability	3 [0 to 5]	3 [0 to 5]
Bladder disability	0 [0 to 5]	2 [0 to 5]
Bowel disability	0 [0 to 4]	0 [0 to 4]
Fatigue	3 [0 to 4]	2 [0 to 4]
Sexual disability	0 [0 to 5]	0 [0 to 5]
Other disabilities	1 [0 to 4]	1 [0 to 4]
GNDS sum score	12 [0 to 28]	13 [0 to 33]

Inter-rater agreement on the different sub-scales was variable, with kappa coefficients ranging between 0.24 and 0.72 (fair to substantial), intraclass correlation coefficients ranging between 0.46 and 0.91 (moderate to almost perfect), and repeatability coefficients ranging between 0.4 and 2.9 point. Inter-rater agreement on the GNDS-R sum score was relatively high with kappa coefficient of 0.12 (fair), an intraclass correlation coefficient of 0.91 (almost perfect), and a repeatability coefficient of 11.1 point (Table 6.18). The mean score differences between the two raters for many GNDS-R sub-scales, but not the GNDS-R sum score, were slightly biased towards the patient's relatives suggesting that they tended to over estimate patient's disabilities.

The mean (SD) of the time needed for scale administration and scoring was 4 minutes and 37 seconds (2 minutes and 36 seconds) for the neurologist (myself), and 7 minutes and 37 seconds (6 minutes and 25 seconds) for the patients' relatives.

Table 6.18 GNDS-R doctor-relative inter-rater reliability ($n = 39$)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Cognitive disability	0.43 [0.24 to 0.62]	0.75 [0.06 to 0.99]	-0.31 [-0.64 to 0.03]	2
Mood disability	0.24 [0.04 to 0.44]	0.46 [0.01 to 0.99]	-0.18 [- 0.52 to 0.16]	2.1
Visual disability	0.52 [0.27 to 0.77]	0.76 [0.33 to 0.99]	-0.10 [-0.27 to 0.06]	1
Speech disability	0.37 [0.07 to 0.67]	0.56 [0.03 to 0.99]	- 0.08 [-0.30 to 0.13]	1.4
Swallowing disability	0.72 [0.49 to 0.95]	0.91 [0.70 to 0.99]	- 0.03 [-0.12 to 0.07]	0.4
Upper limb disability	0.66 [0.48 to 0.84]	0.75 [0.19 to 0.99]	- 0.21 [-0.50 to 0.09]	1.8
Lower limb disability	0.60 [0.42 to 0.78]	0.87 [0.22 to 0.99]	0 [-0.19 to 0.19]	1.6
Bladder disability	0.58 [0.38 to 0.78]	0.79 [0.20 to 0.99]	-0.18 [-0.49 to 0.09]	1.8
Bowel disability	0.75 [0.57 to 0.93]	0.88 [0.25 to 0.99]	- 0.10 [-0.32 to 0.12]	1.3
Fatigue disability	0.49 [0.30 to 0.68]	0.65 [0.05 to 0.99]	- 0.23 [-0.71 to 0.25]	2.9
Sexual disability	0.63 [0.40 to 0.86]	0.72 [0.17 to 0.99]	- 0.26 [-0.71 to 0.20]	2.8
Other disabilities	0.48 [0.18 to 0.68]	0.65 [0.12 to 0.99]	- 0.15 [-0.49 to 0.18]	2.
GNDS sum score	0.12 [0.00 to 0.24]	0.91 [0.21 to 0.99]	0 [-1.28 to 1.28]	11.1

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

These results suggest that the GNDS-R and many of its sub-scales are reliable when administered by a non-neurologist health care personnel, or by patients' relatives or carers.

6.5.6 The reliability / validity of the GNDS-R when administered over the telephone or via a postal questionnaire

The GNDS–R was originally designed for interview administration. However the feasibility of administering this scale over the telephone or via a postal questionnaire was tested since such applications could simplify the conduct of clinical trials and reduce their costs.

I contacted all the patients who took part in the intra-rater reliability and responsiveness study by telephone within three days of their third three monthly visits and re-administered GNDS-R over the telephone. Records of the time needed to administer the scale were kept during these assessments. A postal

version of the GNDS-R was devised using the exact wording of the interview sections with a slightly modified layout to simplify their readability. This 'postal questionnaire' was piloted on a group of five naive patients to assess its intelligibility and readability, and the layout, but not the questions, was modified accordingly (Appendix 2). Copies of this questionnaire were given to all patients who took part in the intra-rater reliability study during their fourth three monthly visits, and the patients were asked to complete the questionnaires and return them within three days. They were also asked to keep a record of the time needed to complete these questionnaires. All returned questionnaires were later scored by myself according to the patients' responses.

A). Interview vs. telephone administration

Forty-seven patients took part in this study. I was unable to contact the other three patients within the specified time frame. The median (range) of the GNDS-R sum and sub-scores of the interview and the telephone administrations were almost identical (Table 6.19). With the exemption of the mood disability sub-scale, the correlations between the scores obtained by administering the scale during an interview or over the telephone by the same rater were high with correlations coefficients ranging between 0.84 and 1.

Table 6.19 Median (range) GNDS-R scores obtained during interview and telephone scale administration ($n = 47$)

Scale item	Interview application Median (range)	Phone application Median (range)	*r*
Cognitive disability	0 [0 to 4]	0 [0 to 4]	0.86
Mood disability	0 [0 to 3]	0 [0 to 3]	0.67
Visual disability	0 [0 to 3]	0 [0 to 3]	0.84
Speech disability	0 [0 to 3]	0 [0 to 3]	1
Swallowing disability	0 [0 to 2]	0 [0 to 2]	0.93
Upper limb disability	1 [0 to 4]	1 [0 to 4]	0.95
Lower limb disability	2 [0 to 5]	2 [0 to 5]	0.99
Bladder disability	0 [0 to 5]	0 [0 to 5]	0.92
Bowel disability	0 [0 to 5]	0 [0 to 5]	0.93
Fatigue	2 [0 to 4]	2 [0 to 4]	0.91
Sexual disability	0 [0 to 5]	0 [0 to 5]	1
Other disabilities	1 [0 to 4]	0 [0 to 4]	0.84
GNDS sum score	12 [0 to 28]	12 [0 to 29]	0.96

Inter-rater agreement on the different sub-scales was high, with kappa coefficients ranging between 0.69 and 1 (substantial to almost perfect), intraclass correlation coefficients ranging between 0.70 and 1 (substantial to almost perfect), and repeatability coefficients ranging between 0 and 1.4 points. Inter-rater agreement on the GNDS-R sum scores was equally high with intraclass correlation coefficient of 0.96 (almost perfect), and a repeatability coefficient of 4.1 points (Table 6.20). The mean score differences between the two scale applications were small with relatively narrow 95% confidence intervals which included the '0' value indicating the absence of any raters' bias. The mean (SD) of the time needed for scale administration and scoring was 5 minutes and 12 seconds (3 minutes and 50 second) for the interview application, and 5 minutes and 1 seconds (2 minutes and 30 seconds) for the telephone application.

Table 6.20 GNDS-R interview / telephone scale administration inter-rater reliability (n = 47)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Mental disability	0.81 [0.64 to 0.98]	0.85 [0.54 to 0.99]	0.09 [-0.13 to 0.30]	1.4
Mood disability	0.69 [0.48 to 0.90]	0.70 [0.39 to 0.99]	0 [-0.21to 0.21]	1.4
Visual disability	0.76 [0.59 to 0.93]	0.87 [0.58 to 0.99]	-0.04 [-0.15 to 0.06]	0.7
Speech disability	1 N/A	1 N/A	0 [N/A]	0
Swallowing disability	0.79 [0.59 to 0.99]	0.87 [0.59 to 0.99]	-0.09 [-0.19 to 0.02]	0.7
Arm disability	0.86 [0.74 to 0.98]	0.97 [0.80 to 0.99]	0.11 [0.01 to 0.19]	0.6
Leg disability	0.97 [0.95 to 0.99]	0.99 [0.84 to 0.99]	0.02 [-0.02 to 0.06]	0.3
Bladder disability	0.77 [0.63 to 0.91]	0.91 [0.37 to 0.99]	0.13 [-0.04 to 0.19]	1.1
Bowel disability	0.91 [0.84 to 0.98]	0.96 [0.47 to 0.99]	-0.08 [-0.26 to 0.08]	1.1
Fatigue	0.79 [0.65 to 0.93]	0.92 [0.43 to 0.99]	0.11 [-0.11 to 0.32]	1.4
Sexual disability	1 N/A	1 N/A	0 [N/A]	0
Other disabilities	0.77 [0.62 to 0.92]	0.83 [0.45 to 0.99]	0 [-0.15 to 0.15]	1.4
GNDS sum score	0.26 [0.09 to 0.42]	0.96 [0.54 to 0.99]	0 [-0.43 to 0.43]	4.1

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

B). Interview vs. postal questionnaire scale administration

Forty-two patients took part in this study. Two patients did not complete the questionnaire in the specified time, and six questionnaires were lost in the post. The median (range) of the GNDS-R sum and sub-scores of the interview and the postal questionnaire scale administration were almost identical. The correlations between the scores obtained by administering the scale during an interview or via a postal questionnaire were high for the GNDS-R sum score and its speech, swallowing, upper limb, lower limb, bladder, bowel, fatigue, and sexual function disability sub-scales with correlations coefficients ranging between 0.81 and 1, and moderate for the cognition, mood, visual, and other disabilities sub-scales with correlations coefficients ranging between 0.66 and 0.76 (Table 6.21).

Table 6.21 Median (range) GNDS-R scores obtained during interview and postal questionnaire administration ($n = 42$)

Scale items	Interview application Median (range)	Postal questionnaire Median (range)	*r*
Mental disability	0 [0 to 4]	0 [0 to 4]	0.66
Mood disability	0 [0 to 3]	0 [0 to 3]	0.65
Visual disability	0 [0 to 3]	0 [0 to 3]	0.76
Speech disability	0 [0 to 3]	0 [0 to 3]	0.89
Swallowing disability	0 [0 to 2]	0 [0 to 2]	0.94
Upper limb disability	1 [0 to 4]	1 [0 to 4]	0.91
Lower limb disability	2 [0 to 5]	2 [0 to 5]	0.92
Bladder disability	0 [0 to 5]	1 [0 to 5]	0.86
Bowel disability	0 [0 to 4]	0 [0 to 4]	0.93
Fatigue	2 [0 to 4]	2 [0 to 4]	0.81
Sexual disability	0 [0 to 5]	0 [0 to 5]	0.93
Other disabilities	1 [0 to 3]	2 [0 to 3]	0.70
GNDS sum score	12 [0 to 28]	13 [0 to 29]	0.93

The reliability of the different GNDS-R sub-scales using the two scale administration methods was high, with kappa coefficients ranging between 0.49 and 0.84 (moderate to almost perfect), intraclass correlation coefficients ranging between 0.761 and 0.92 (substantial to almost perfect), and repeatability

coefficients ranging between 0.5 and 1.7 points. The reliability of the GNDS-R sum score was equally high with an intraclass correlation coefficient of 0.93 (almost perfect), and a repeatability coefficient of 6.2 points (Table 6.22). The mean score differences between the two scale applications were small with relatively narrow 95% confidence intervals which included the '0' value indicating the absence of any bias. The mean (SD) of the time needed for scale administration and scoring was 5 minutes and 23 seconds (3 minutes and 32 second) for the interview application, and 6 minutes and 30 seconds (3 minutes and 54 seconds) for the postal questionnaire application (excluding the time needed for scoring the questionnaire).

Table 6.22 GNDS-R interview-postal questionnaire scale administration inter-rater reliability ($n = 42$)

Scale item	Kappa (95% CI)	ICC (95% CI)	Mean (95% CI) score difference	RC
Cognitive disability	0.49 [0.27 to 0.77]	0.61 [0.12 to 0.99]	-0.05 [-0.54 to 0.36]	2.6
Mood disability	0.55 [0.31 to 0.79]	0.69 [0.18 to 0.99]	-0.02 [-0.26 to 0.21]	1.5
Visual disability	0.70 [0.51 to 0.89]	0.77 [0.35 to 0.99]	-0.14 [-0.29 to 0.04]	0.5
Speech disability	0.64 [0.53 to 0.75]	0.88 [0.60 to 0.99]	-0.05 [-0.17 to 0.07]	0.7
Swallowing disability	0.79 [0.59 to 0.99]	0.92 [0.73 to 0.99]	-0.02 [-0.11 to 0.06]	0.5
Upper limb disability	0.81 [0.67 to 0.95]	0.93 [0.77 to 0.99]	-0.02 [-0.12 to 0.17]	0.9
Lower limb disability	0.76 [0.61 to 0.91]	0.92 [0.73 to 0.99]	0.17 [-0.04 to 0.37]	0.9
Bladder disability	0.64 [0.46 to 0.82]	0.87 [0.58 to 0.99]	0 [-0.24 to 0.24]	1.2
Bowel disability	0.79 [0.62 to 0.95]	0.93 [0.75 to 0.99]	-0.02 [-0.19 to 0.14]	1
Fatigue	0.55 [0.37 to 0.73]	0.83 [0.48 to 0.99]	0.17 [-0.16 to 0.49]	0.8
Sexual disability	0.84 [0.67 to 0.99]	0.92 [0.72 to 0.99]	-0.07 [-0.29 to 0.15]	1.4
Other disabilities	0.56 [0.36 to 0.76]	0.69 [0.17 to 0.99]	0 [-0.28 to 0.28]	1.7
GNDS sum score	0.21 [0.02 to 0.40]	0.93 [0.75 to 0.99]	0.02 [-1.02 to 0.97]	6.3

ICC = intraclass correlation coefficient; RC = repeatability coefficient; CI = confidence intervals.

These results suggest that the GNDS-R and many of its sub-scales are reliable when applied over the telephone or via a postal questionnaire. They also

suggest that the telephone and the postal questionnaire applications are valid methods of administering the GNDS-R.

6.6 GNDS-R postal survey

To evaluate the performance of the GNDS-R in a large community-based cohort of patients, and to assess its acceptability to patients, I conducted a study to test the postal version of the scale in a large cohort of naive patients.

6.6.1 Study design

Copies of the GNDS-R postal questionnaire (Appendix 2) were posted to a group of 400 randomly selected patients with multiple sclerosis whose names and addresses were retrieved from the Multiple Sclerosis Resource Centre (a charitable organisation) mailing list. The patients were invited to complete the questionnaires and to provide simple demographic information related to their age, sex, work status, and to indicate whether or not they required help to complete these questions. Participants were also asked to complete a face validity questionnaire identical to that used in the face and content validity studies discussed earlier. A single reminder was posted 6 weeks after the first mailing. All returned questionnaires were scored by a single rater. Uncompleted or incorrectly completed sub-scales were treated as missing values. Missing values were interpolated (mean of the other available scores) if >80% of the other sub-scores were available.

6.6.2 Results

One hundred and ninety four patients (48.5%) completed and returned their questionnaires. The median age of the respondents was 49 years (range 27-73), 78% were female, 12% were in full time and 15% were in part time employment. Eighty percent of the respondents were able to complete the questionnaires by themselves, whereas the rest needed some help in reading the questions (7%), working out the answers (8%), or writing them down (15%) (Table 6.23)

Table 6.23 Demographic characteristics of the GNDS-R postal survey respondents

Number of patients surveyed	400
Number of respondents	194
Return rate	48.5 %
Age: median (range) years	49 (27 to 73)
Sex: female / male	151 / 43
Work status:	
Full time	23 (11.9 %)
Part time	29 (14.9 %)
Not working	142 (73.2 %)
Help needed in filling in the questionnaire	53 (27.3 %)
Type of help required	
Reading the questions	14 (7.2 %)
Writing in the answers	29 (14.9 %)
Deciding on how to answer the questions	15 (7.7 %)

Eighty-three percent of the respondents completed the GNDS-R questionnaire fully and correctly. The other 17% left parts of the questionnaires unanswered or incorrectly completed (Table 6.24). After interpolating for missing values, sum scores could be assigned to 97.9% of the returned questionnaires. The frequency of the GNDS-R sum scores was normally distributed with no 'ceiling' or 'floor' effects (Figure 6.3).

Near normal frequency distribution was also observed in the cognitive, mood, upper limb, lower limb, bladder, bowel, fatigue, sexual function, and other disability sub-scales, whereas a skewed distribution to the 'less disabled' end of the scale was noted in the vision, speech, and swallowing sub-scales reflecting the previously noted mild severity prevalence figures of these disabilities in population-based cohorts (Rodriguez et al., 1994; Midgard et al., 1996) (Table 6.24).

Figure 6.3 The frequency distribution (with the superimposed normal curve) of the GNDS-R postal questionnaire (n = 190)

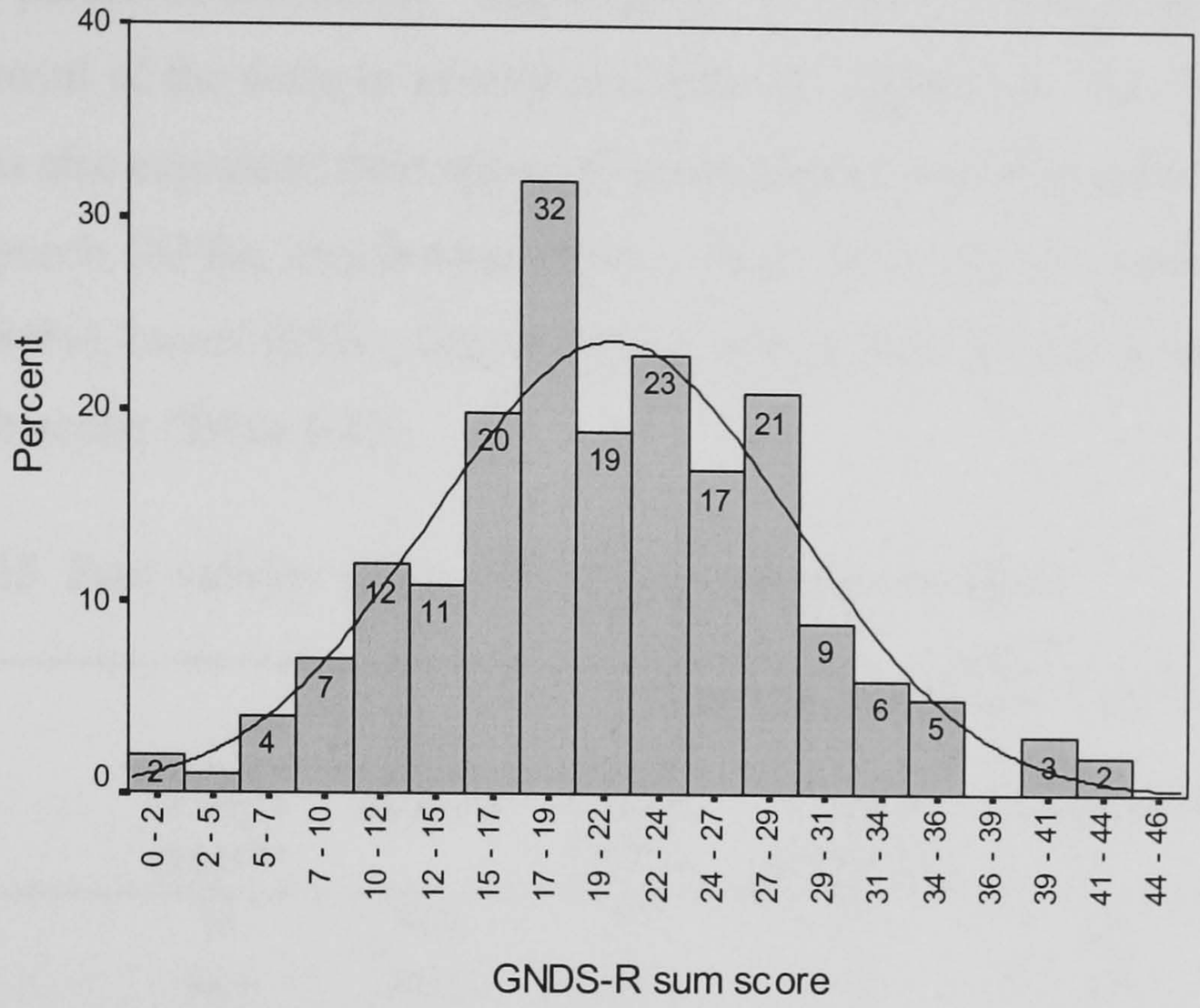


Table 6.24 Frequency distribution of the GNDS-R postal questionnaire (% of respondents scoring at each individual grade, n = 194)

	Grades						
Scale item	0	1	2	3	4	5	Missing values (n)
Cognitive disability	36.1	12.3	39.2	5.7	6.7	0	2
Mood disability	27.3	28.4	12.9	26.3	4.6	0.5	2
Visual disability	37.1	51.6	7.2	2.6	0	1.5	4
Speech disability	73.6	13.9	8.2	3.6	0.5	0	7
Swallowing disability	77.9	8.2	10.3	3.1	0	0.5	4
Upper limb disability	27.8	5.7	25.3	30.9	7.2	3.1	1
Lower limb disability	6.2	22.7	26.8	4.7	27.3	12.4	4
Bladder disability	14.9	2.1	34.5	24.2	12.9	11.3	0
Bowel disability	40.2	7.3	23.2	2.6	24.2	2.6	4
Fatigue disability	2.6	4.6	15.4	49.5	6.7	17.5	1
Sexual disability	35.1	5.2	8.8	7.7	30.9	12.4	17
Other disabilities	13.9	10.3	25.8	27.9	39.2	8.8	4

The majority of the respondents (81.4%) completed the face validity questionnaire, whereas the rest (18.6%) left part of the questionnaire unanswered. Many respondents found the GNDS-R questionnaire not to be comprehensive enough and the 'yes/no' answer options to be restrictive but the scale was otherwise perceived favourably. The majority (87.7%) of the respondents expressed their approval of the scale in general and only 4% disliked it. The majority of the responders also expressed their approval of the cognitive (87%), mood (83%), vision (84%), speech (83%), swallowing (84%), upper limb (82%), lower limb (84%), bladder (86%), bowel (85%), fatigue (86%), sexual function (89%) and the 'others' (77%) sub-scales (Table 6.25).

Table 6.25 Face validity of the GNDS-R postal questionnaire

	% of respondents (n = 194)						
Scale item	Strongly approve	Approve	Tends to approve	Tends to disapprove	Disapprove	Strongly disapprove	Missing values
Cognition	33	46.9	7.2	1.5	2.6	0	8.8
Mood	34.4	41.2	7.2	4.6	2.1	0.5	10
Vision	40.7	36.1	6.7	3.1	3.1	0	11
Speech	33.5	42.3	7.2	3.6	2.1	0	11.3
Swallowing	34	40.2	9.8	2.1	0.5	0.5	12.9
Upper limb	38.1	34	10.3	3.1	2.6	2	9.9
Lower limb	42.3	31.4	9.8	2.6	1	2.6	10.3
Bladder	42.3	35.1	8.2	2.6	1.5	0.5	9.8
Bowel	38.1	36.6	9.8	1.5	1.5	0.5	12
Fatigue	46.4	32	7.2	2.1	2.1	0.5	9.7
Sexual function	33.5	34.5	19.8	11	1	6.1	5.9
Others	32.5	36.6	7.7	3.1	1	0.5	18.6
Overall scale	45.4	33.5	8.8	1.5	1.5	0	9.3

Although the various GNDS-R items were not directly drawn from in-depth patient interviews (Thompson and Hobart, 1996b), these results support the face validity of the scale by indicating patients' approval of the scale and its various items (Sharrack and Hughes, 1999a).

6.7 Factor analysis

Factor analysis was conducted to assess the underlying conceptual dimensions of the 12 GNDS-R sub-scales and investigate the presence of any redundant items. Data obtained during the postal survey met all the criteria required for an adequate factor analysis and these results will be discussed. Factor extraction by principal components, followed by orthogonal or oblique rotation gave comparable results. Factors which gained an eigenvalue of <1, and items which gained a loading of <0.35 on any factor so created were discarded. This analysis suggested a four factor solution which accounted for 58.7% of the total variance (cumulative percentage of 26.5%, 39.9%, 49.9%, and 58.7%; eigenvalues of 3.18, 1.61, 1.25, 1.05 respectively). The rotated matrix suggested no redundant items. With a cut off loading value of 0.35, each GNDS-R sub-scale loaded on one factor only.

The first factor of the rotated matrix (*spinal factor)* correlated with the lower limb, bladder, bowel, and sexual function sub-scales. The second factor (*mental factor)* correlated with the cognition, mood, and fatigue sub-scales. The third factor (*bulbar factor*) correlated with the speech and swallowing sub-scales. The fourth factor (*upper limb / vision / other disabilities factor*) correlated with the upper limb, vision, and other disabilities sub-scales (Table 6.26). These four factors segregated the 12 GNDS-R dimensions in a logical and predictable manner therefore strengthening the content validity of the scale.

Factor analysis was also conducted on the inter-rater reliability data set with comparable results. The analysis suggested a four factor solution which accounted for 67.8% of the total variance (cumulative percentage of 31.7%, 45.6%, 59.1%, and 67.8%; eigenvalues of 3.81, 1.67, 1.61, 1.05 respectively). Compared with the factor analysis of the postal survey data set, the various GNDS-R sub-scales loaded on the four factors in an identical manner, although the individual contributions of the mental, mood and the upper limb/vision/other to the total variance were different in this sample (Table 6.27).

Table 6.26 GNDS-R factor analysis (postal survey data set): rotated components matrix

Scale item	Factor 1	Factor 2	Factor 3	Factor 4
Cognitive disability	0.09	0.23	**0.75**	0.03
Mood disability	0.30	-0.09	**0.75**	0.17
Visual disability	- 0.09	0.29	0.05	**0.55**
Speech disability	0.07	**0.82**	0.21	-0.01
Swallowing disability	0.12	**0.77**	-0.01	0.27
Upper limb disability	0.29	0.35	0.12	**0.66**
Lower limb disability	**0.61**	0.08	-0.21	0.31
Bladder disability	**0.81**	-0.06	-0.03	-0.03
Bowel disability	**0.65**	0.15	0.12	-0.02
Fatigue	-0.01	0.29	**0.56**	0.25
Sexual disability	**0.62**	0.16	0.32	-0.05
Other disabilities	0.06	-0.12	0.21	**0.79**

Table 6.27 GNDS-R factor analysis (inter-rater reliability data set): rotated components matrix

Scale item	Factor 1	Factor 2	Factor 3	Factor 4
Cognitive disability	-0.03	0.17	0.34	**0.57**
Mood disability	0.02	0.08	0.13	**0.79**
Visual disability	0.01	**0.72**	0.21	0.06
Speech disability	0.21	0.09	**0.74**	0.19
Swallowing disability	0.04	0.02	**0.88**	-0.05
Upper limb disability	0.25	**0.71**	0.22	0.31
Lower limb disability	**0.74**	0.35	0.19	0.07
Bladder disability	**0.81**	0.12	0.27	0.07
Bowel disability	**0.81**	-0.26	-0.04	-0.25
Fatigue	0.21	0.26	-0.05	**0.77**
Sexual disability	**0.67**	0.34	-0.02	-0.02
Other disabilities	0.19	**0.72**	-0.03	0.29

6.8 Validity

In the absence of a gold standard to assess disability in multiple sclerosis, the validity of GNDS-R was established through the process of construct validity by assessing the degree to which this scale correlated with other existing measures of disability (convergent construct validity), impairment, handicap, and health related quality of life measures (discriminant construct validity), and other generated hypotheses (construct validity by hypothesis testing), and by testing its ability to differentiate between patient groups known to differ in the degree of their disability (construct validity by group differences).

6.8.1 Study design

The validity study was conducted on the same cohort of 50 patients who took part in the intra-rater and responsiveness study described above. During their third three monthly visit all patients were additionally assessed as follows:

A). All patients underwent full neurological examinations including assessment of their visual acuity using a Snellen Chart, and were assigned scores on the SNRS, EDSS, FIM, AI, CMBS, time to walk 10 metres, and the Barthel Index.

B). Patients were also asked to complete 12 separate visual analogue scales, adopted from the EuroQol VAS, to indicate the degree of disability as perceived by them in each of the 12 GNDS-R dimensions.

C). Upper limb function was assessed using the 9-hole peg test (time taken to place 9 pegs into a board and remove them) as an impairment measure (Mathiowetz et al., 1985). Three trials with each hand were performed and the mean of the results with the two hands was calculated.

D). Swallowing was assessed using the timed swallowing test devised by Nathadwarawala and co-workers as an impairment test (Nathadwarawala et al., 1992). The nature of the test was explained to the patients who were given 150 ml of cold water to drink from a standard cup. Patients were asked to drink the water as quickly as possible but to take care and to stop if difficulty arose. During the procedure, I sat at the side of the subjects to obtain an adequate view of the laryngeal movements during swallowing. The time taken from the beginning of the test to the last swallow recognised by return of the larynx to the rest position

was noted. The residual volume was measured in those who were unable to complete the test and the swallowing speed (ml/s) was calculated.

E). Speech was assessed using the Frenchay Dysarthria assessment as an impairment measure (Enderby, 1988). This test was administered by giving patients specific tasks designed to assess the anatomical, physiological, and perceptual features of dysarthria in 8 separate sections: reflexes (3 items), respiration (2 items), lip movements (5 items), jaw movements (2 items), palatal movements (3 items), phonation (5 items), tongue movements (6 items), and intelligibility (3 items). A nine-point scoring system was used to record the patient's response in each sub-set, and the results were charted on a bar graph with a nine-point scale on the vertical axis, with eight sets and corresponding subsets on the horizontal axis giving a profile of the speech assessments. Sum scores of the eight sets were also calculated.

F). All patients underwent a detailed battery of cognitive tests which were constructed following consultation with a neuropsychologist from the National Hospital for Neurology and Neurosurgery (Dr Luke Kartsounis). This battery was constructed as a comprehensive battery of impairment measures which included:

1). The National Adult Reading Test (NART) as a measure of pre-morbid level of functioning (Nelson and Willison, 1992).

This test consists of 50 single irregular words which do not obey the usual English letter-to-sound rules. The raw scores were transformed into IQ equivalents to provide an estimate of pre-morbid intelligence.

2). Letter cancellation task as a measure of visual attention (Willison et al., 1980).

This task was devised and used previously in a clinical study to assess the effect of high haematocrit on alertness. Patients were asked to stroke out at speed all the B's from a random array of 5 capital letters (A, B, C, D and E) arranged in a matrix of 11 rows by 8 columns on an A4 sheet. The time to complete the task and the number of errors were recorded.

3). Paced Auditory Serial Addition Task (PASAT) version 2 and 4 seconds as a measure of auditory attention and speed of information processing (Gronwall, 1977).

Two pre-recorded lists of 61 numbers read at a rate of one digit every 2 or 4 second intervals were played, and the patients were asked to add each number to the previous one and give their answers aloud. A demonstration with written

numbers and two practice auditory lists of 10 digits recorded at a rate of 2 and 4 second intervals were given before the test was conducted. The total of the correct answers at each paced speed (ranging between 0 and 60) was recorded.

4). Standard Progressive Matrices as a measure of non-verbal abstract reasoning and general intellectual function (Raven, 1994).

The test consists of 60 problem (diagrammatic puzzles) divided into five sets (A, B, C, D, and E) each made up of 12 problems. In each set the first problem is self-evident, but the problems which follow build on the argument of those that have gone before and become progressively more difficult. A self-administered version was used. The test was explained to the patients who were asked to complete the task at their own pace up to a maximum of one hour. The final score was the total number of problems solved correctly. The consistency of the answers were checked as described in the operational manual, and the scores were transformed into percentiles and grades ranging between (I) indicating intellectual superiority, and (IV) indicating intellectual impairment.

5). Similarities sub-set of the Wechsler Adult Intelligence Scale – Revised as a measure of verbal abstract reasoning (Wechsler, 1981).

Patients were presented with a list of 14 two-item sets and were asked to indicate in what way the two items were alike. Raw scores were transformed into standard scores as described in the operational manual.

6). Immediate and delayed story recall sub-set of the Adult Memory and Information Processing Battery as a measure of verbal memory (Coughlan and Hollows, 1985).

Patients were read a standard story and were asked to repeat it immediately and 30 minutes afterwards. Raw scores were transformed into standard scores using the operational manual.

7). Controlled Oral Word Association Test (COWAT) as a measure of verbal fluency and planning abilities (Benton and Hamsher, 1976; Miller, 1984).

Patients were asked to say as many words as they could think of in 60 seconds that begin with a given letter of the alphabet (F, A, S) excluding proper nouns, numbers, and the same word with a different suffix. A practice trial using the letter C was administered first to assure that the patient comprehended the task. The raw scores, the sum of all acceptable words produced in the three one-

minute trials, were adjusted for age, sex, and education and the adjusted scores were converted into percentiles.

F). All patients were provided with booklets containing copies of the following self-administered scales:

1). The London Handicap Scale (Harwood et al., 1994).

This self-administered scale has six items each consisting of a question related to one of the ICIDH handicap dimensions (mobility, orientation, physical independence, occupation, social integration, and economic self-sufficiency), with a six-point scoring system from 1 to 6. Patients' responses are weighted and summed giving a total score ranging between 0 and 100.

2). The Short Form 36 health survey questionnaire (SF-36) (Garratt et al., 1993).

This is a 36 item self-administered questionnaire which addresses eight distinct quality of life dimensions including physical functioning (10 items), role limitation due to physical problems (4 items), role limitation due to emotional problems (3 items), social functioning (2 items), mental health (5 items), energy / vitality (4 items), pain (2 items), and general health perception (5 items), with a further single unscaled item related to health change over the last one year. Each item comprises a question requesting information about a certain aspect of perceived health. For each dimension, item scores are coded, summed, and transformed into a scale from 0 (worst possible health status) to 100 (best possible health status).

3). The General Health Questionnaire (GHQ) (Goldberg and Williams, 1988).

This is a self-administered screening test was designed to detect non-psychotic psychiatric symptoms among respondents in community settings or general medical outpatient departments, and can be considered to be an impairment measure. The 28-item scaled version is used mainly for research proposes. This questionnaire has four scales testing somatic symptoms, anxiety and insomnia, social dysfunction, and severe depression. Each scale has seven items which consist of questions asking whether the participant has recently experienced a particular symptom or certain behaviour. The answers were scored on the 4-point 'Likert' scale between 0 (less than usual) and 3 (much more than usual) rather than the bimodal response (0-0-1-1) since the former is thought to produce less skewed distribution and offer marginal advantages if sub-scale scores were required.

4). The Beck Hopelessness Scale (Beck et al., 1974)

This is a 20 true-false item self-administered scale which measures the extent of negative expectations about the immediate and long-range future (pessimism) as perceived by the patient, and is considered to be an impairment measure. The items scored are summed to yield a total score that can range from 0 to 20 with the higher scores indicating greater hopelessness.

5). The Chalder and Fatigue Scale (Chalder et al., 1993)

This is a self-administered 14-item scale which was developed to measure the severity of physical and mental fatigue symptoms, and as such it is considered as an impairment scale. The scale items are answered on a 4-point scale and scored using the General Health Questionnaire bimodal response method (0-0-1-1) giving a total fatigue score which ranges from 0 to 28. The scale can also provide separate mental (ranging between 0 and 12) and physical (ranging between 0 and 16) fatigue scores.

6). The Golombock Rust Inventory of Sexual Satisfaction (GRISS) (Rust and Golombok, 1985)

This is a self-administered questionnaire which was developed to provide objective assessments of the quality of sexual relationships and the function of individuals within them, and can be considered as a sexual disability scale. It has 28 items on a single sheet all answered on a 5-point scale. The scale gives a transformed overall sexual dysfunction score on a 10 point scale between 0 and 9 with scores 5 and above indicating a problem. It also gives a sexual profile using similar 10-point scales on 12 sub-scales comprising impotence and premature ejaculation in males, anorgasmia and vaginismus in females, infrequency, non-communication, non-sensuality, avoidance, and dissatisfaction of both sexes.

Patients were asked to compete these 6 questionnaires and return them within three days of their third three monthly clinical assessment.

H). All patients were also ranked by myself according to their ability to work, do their housework, and look after themselves. They were also ranked independently by myself and a research nurse according to their subjectively perceived degree of disability.

I). Patients' close relatives or carers were asked to complete a copy of the GNDS-R questionnaire to reflect their independent assessment of the patients'

various disabilities as they observed or perceived them without referring to the patients.

Convergent and discriminant construct validity were tested by assessing the degree to which the GNDS-R and its 12 sub-scales correlated with other measures of impairment (SNRS, EDSS, cognitive assessment, General Health Questionnaire, Beck Depression Inventory, visual acuity, Frenchay Dysarthria Assessment, swallowing test, 9-hole peg test, time to walk 10 metres, and Chalder Fatigue Scale), disability (EDSS, FIM, Barthel Index, the disability domain of the CAMBS, the various visual analogue scales, GRISS sexual dysfunction scale), handicap (the London Handicap Scale, the handicap domain of the CAMBS), and health related quality of life (EuroQol, and SF-36). Group differences construct validity was assessed by testing the extent to which the GNDS-R scores correlated with the severity of disability as judged by the two raters. Hypothesis testing construct validity was assessed by testing the hypothesis that GNDS-R scores should be more abnormal in patients who are unable to work or do their housework because of multiple sclerosis, and in patients who are dependent on others for some or all of their activities of daily living. The objectivity of the GNDS-R was also assessed by testing the hypothesis that GNDS-R scores should correlate with scores obtained when the scale is completed by patients' close relatives or carers independently of patients' own assessments of their disability.

As indicated above, 50 patients were included in this study. The group consisted of 31 women and 19 men with a median age 36 years (range 24–51), and a median disease duration of 12 years (range 2-17). The patients had mild to moderate disability with a median EDSS score of 4.5 (range 0-7.5), and median (range) GNDS-R score was 12 (0-28).

6.8.2 Convergent and discriminant validity

A). Cognitive disability sub-scale

The distribution of pre-morbid IQ, estimated from the NART, showed a mean of 108 and a standard deviation of 13 indicating that the patients constituted a relatively unbiased sample from the general population (a perfect normal sample would have a mean of 100 and a standard deviation of 15). One patient was not able to perform the letter cancellation task due to severe tremor of both arms. Two other patients claimed that their performance was affected by loss of dexterity. Statistical analysis with or without their data revealed similar results.

The GNDS-R cognitive disability sub-scale correlated weakly with the speed of letter cancellation (r = 0.31), and the mentation and mood domain of the SNRS (r = -0.54), and moderately with the mental Functional System of the EDSS (r = 0.62) but not with any other impairment measure. It also correlated highly with the memory domain of the FIM (r = -0.80), moderately with perceived degree of cognitive disability (r = -0.70), and weakly with the problem solving domains of the FIM (r = -0.33). It also correlated weakly with the orientation domain of the London Handicap scale (r = -0.36), but not with any aspects of Health Related Quality of life (Table 6.28).

As the strongest correlations were established with other disability rather than impairment, handicap, or heath related quality of life measures, these results support the validity of the cognitive disability sub-scale as a measure of disability.

Table 6.28 The correlation between the GNDS-R cognitive disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Cognitive disability sub-scale *	0 (0 to 3)	-	-
Pre-morbid IQ (NART) **	108 (13)	-	-
Impairment measures			
Speed of letter cancellation *(seconds)* ** #	23.7 (7.8)	0.31	0.027
PASAT 2" (correct answers) *	34 (4 to 59)	-0.09	NS
PASAT 4" (correct answers) *	51 (20 to 60)	-0.26	0.057
Standard progressive matrices (grades) *	3 (0 to 5)	0.13	NS
Similarities (standard scores) *	10 (6 to 15)	0.08	NS
Immediate story recall (percentiles) **	58 (27)	0.18	NS
Delayed story recall (percentiles) **	56 (27)	0.20	NS
COWAT (adjusted scores) *	42 (15 to 62)	0.03	NS
SNRS - Mentation and Mood *	10 (4 to 10)	-0.54	0.002
EDSS - Mental Functional System *	0 (0 to 3)	0.62	<0.001
Disability measures			
FIM – Problem solving *	7 (3 to 7)	-0.33	0.047
FIM – Memory *	7 (0 to 7)	-0.80	<0.001
Perceived cognitive disability (VAS) **	86 (20)	-0.70	<0.001
Handicap measures			
London Handicap Scale – Orientation **	1.4 (1.3)	-0.36	0.013
Health Related Quality of Life measures			
SF 36 – Mental health **	72 (9)	-0.12	NS

* Median (range); ** Mean (SD); # n = 47 patients

B). Mood disability sub-scale

The GNDS-R mood disability sub-scale correlated moderately with the mentation and mood item of the SNRS (r = -0.59), and the mental Functional System of the EDSS (0.40), and weakly with overall score of the General Health Questionnaire (r = 0.34), but not with any of its four domains, nor with the Beck Hopelessness Scale. It also correlated moderately with the perceived degree of mood disability (r = -0.74), and weakly with the emotional role limitation of the

SF-36 (r = -0.45). No handicap measures were available for this study (Table 6.29).

Table 6.29 The correlation between the GNDS-R mood disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Mood disability sub-scale *	0 (0 to 4)	-	-
Impairment measures			
SNRS – Mentation and Mood *	10 (4 to 10)	-0.59	<0.001
EDSS – Mental Functional System *	0 (0 to 3)	0.40	<0.001
General Health Questionnaire *			
GHQ - overall score	22 (19 to 32)	0.34	0.02
GHQ - Somatic symptoms	1 (0 to 5)	0.11	NS
GHQ - Anxiety and insomnia	0 (0 to 4)	0.20	NS
GHQ - Social dysfunction	1 (0 to 5)	0.06	NS
GHQ - Severe depression	7 (7 to 10)	0.19	NS
Beck Depression Scale *	6 (6 to 16)	0.37	NS
Disability Measures			
Perceived mood disability (VAS) **	82 (19)	-0.74	0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
SF 36 – Emotional Role Limitation **	50 (42)	-0.45	0.001

* Median (range); ** Mean (SD)

As the strongest correlations were established with other disability rather than impairment, or heath related quality of life measures, these results support the validity of the mood disability sub-scale as a measure of disability.

C). Visual disability sub-scale

The GNDS-R visual disability sub-scale correlated moderately with the upper cranial nerves (visual acuity, fields, discs, pupils, eye movements, and nystagmus) components of the SNRS (r = -0.61), the visual EDSS Functional Systems (r = 0.60), and with the visual acuity of the worse (r = 0.51) and the

better eyes (r = 0.49). It also correlated moderately the perceived degree of visual disability (r = -0.79). There were no handicap and health related quality of life measures available for this study (Table 6.30).

Table 6.30 The correlation between the GNDS-R visual disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Visual disability sub-scale *	0 (0 to 3)	-	-
Impairment measures			
SNRS – Upper cranial nerves (items 1 to 4) *	17 (0 to 19)	-0.61	<0.001
EDSS – Visual Functional System *	0 (0 to 6)	0.60	<0.001
Visual acuity - better eye *	6/6 (5/6 to 36/6)	0.49	<0.001
Visual acuity – worse eye *	6/6 (5/6 to 60/6)	0.51	<0.001
Disability Measures			
Perceived visual disability (VAS) **	83 (21)	-0.79	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (SD)

As the strongest correlations were established with other disability rather than impairment measures, these results support the validity of the visual disability sub-scale as a measure of disability.

D). Speech disability sub-scale

The GNDS-R speech disability sub-scale correlated moderately with the lower cranial nerves domain of the SNRS (r = -0.72), and weakly with the brain stem Functional System of the EDSS (r = 0.52). It also correlated weakly with the Frenchay Dysarthria Assessment sum score (r = -0.46), and its reflex (r = -0.50), respiration (r = -0.44), lips (r = -0.50), soft palate (r = -0.40), tongue (r = -0.43), and intelligibility (r = -0.58) domains sum scores. The GNDS-R speech disability sub-scale also correlated moderately with the perceived degree speech

disability (r = -0.75). There were no handicap or quality of life measures for comparisons (Table 6.31).

Table 6.31 The correlation between the GNDS-R speech disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Speech disability sub-scale *	0 (0 to 3)	-	-
Impairment measures			
SNRS – Lower cranial nerves *	0 (0 to 5)	-0.72	<0.001
EDSS – Brain stem Functional System *	0 (0 to 4)	0.52	0.002
Frenchay Dysarthria Assessment **			
Sum score	246.6 (10.4)	-0.46	0.001
Reflex	25.6 (2.4)	-0.50	<0.001
Respiration	17.5 (1.2)	-0.44	0.002
Lips	44.5 (1.1)	-0.50	<0.001
Jaw	18 (0)	-0.28	NS
Soft palate	26.8 (0.9)	-0.40	0.004
Larynx	35.4 (1.4)	-0.26	NS
Tongue	52.4 (3.2)	-0.43	0.002
Intelligibility	26.3 (2)	-0.58	<0.001
Disability Measures			
Perceived speech disability (VAS) **	93 (15)	-0.75	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (SD)

The high correlation between this sub-scale and the lower cranial nerves domain of the SNRS is likely to reflect the mild degree of speech disability of this cohort since this sub-scale concentrates on the symptomatic aspects of the speech disturbance in its lower grades. However the high correlation with the perceived degree of speech disability support the validity of this sub-scale as a measure of disability.

E). Swallowing disability sub-scale

The GNDS swallowing disability sub-scale correlated weakly with the lower cranial nerves domain of the SNRS (r = -0.40), and the timed swallowing test (r = 0.33), but moderately with the brain stem Functional System of the EDSS (r = 0.54). It also correlated moderately with the perceived degree of swallowing disability (r = -0.75). There were no other disability, handicap, or quality of life measures available for comparison (Table 6.32).

Table 6.32 The correlation between the GNDS-R swallowing disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Swallowing disability sub-score *	0 (0 to 2)	-	-
Impairment measures			
SNRS – Lower cranial nerves *	0 (0 to 5)	-0.40	0.004
EDSS – Brain stem Functional System *	0 (0 to 4)	0.54	<0.001
Timed swallowing test (Seconds) **	16 (10)	0.33	0.020
Disability Measures			
Perceived disability (VAS) *	94 (11)	-0.75	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (SD)

As the strongest correlation was established with disability rather than impairment measures, these results support the validity of this sub-scale as a measure of disability.

F). Upper limb disability sub-scale

Three patients were unable to perform the 9-hole peg test due to severe tremor or impaired proprioception. The GNDS-R upper limb disability sub-scale correlated moderately with the upper limb items (motor, sensory, cerebellar, and reflexes) of the SNRS (r = -0.53), and the pyramidal Functional System of the EDSS (r = 0.52), but weakly with the cerebellar (r = 0.30) and the sensory (r =

0.41) EDSS Functional Systems, and with the 9-hole peg test time (0.46). It also correlated moderately with the self-care item (items A to E) of the FIM (r = -0.66), the perceived degree of upper limb disability (r = -0.79), the independence domain of the London Handicap Scale (r = -0.63), and with the physical functioning domain of the SF-36 (r = -0.58) (Table 6.33).

Table 6.33 The correlation between the GNDS-R upper limb disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Upper limb disability sub-scale *	1 (0 to 4)	-	-
Impairment measures			
SNRS – Upper limb items (motor, sensory, cerebellar, reflexes) *	24 (6 to 26)	-0.53	<0.001
EDSS – Pyramidal Functional System *	3 (0 to 5)	0.52	<0.001
EDSS – Cerebellar Functional System *	2 (0 to 5)	0.30	0.038
EDSS – Sensory Functional System *	0 (0 to 5)	0.41	0.003
Nine-peg-hole time (mean of right and left hands) (Seconds) ** #	28 (13)	0.46	0.001
Disability Measures			
FIM – Self care items(A to E) *	42 (26 to 42)	-0.66	<0.001
Perceived upper limb disability (VAS) **	90 (17)	-0.76	<0.001
Handicap measures			
London Handicap Scale – Independence **	3 (3.6)	-0.63	<0.000
Health Related Quality of Life			
SF 36 – Physical Functioning **	20 (6)	-0.58	<0.001

* Median (range); ** Mean (SD); # Number = 47 patients

As the strongest correlations were established with other disability rather than impairment, handicap, or heath related quality of life measures, these results support the validity of the upper limb disability sub-scale as a measure of disability.

G). Lower limb disability sub-scale

Six patients were unable to perform the 10-metre walk. The GNDS-R lower limb disability sub-scale correlated highly with the lower limb items (motor, sensory, cerebellar, reflexes, and gait) of the SNRS (r = -0.87), the

pyramidal EDSS Functional System (r = 0.81), and the 10-metre walk (r = 0.87), and moderately with the cerebellar EDSS Functional Systems (r = 0.67), and weakly with sensory EDSS Functional System (r = 0.45). It also correlated moderately with the transfer (items I to K) (r = -0.72) and locomotion (items L and M) (r = -0.73) items of the FIM, and highly with the Ambulation Index (r = 0.90), and the perceived severity of lower limb visual disability (r = -0.80). The lower limb disability sub-scale also correlated weakly with the mobility domain of the London Handicap Scale (r = -0.43), but strongly with the physical functioning domain of the SF-36 (r = -0.87) (Table 6.34).

Table 6.34 The correlation between the GNDS-R lower limb disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Lower limb disability sub-scale *	2 (0 to 4)	-	-
Impairment measures			
SNRS – Lower limb items (motor, sensory, cerebellar, reflexes, gait) *	21 (0 to 39)	-0.87	<0.001
EDSS – Pyramidal Functional System *	3 (0 to 5)	0.81	<0.001
EDSS – Cerebellar Functional System *	2 (0 to 5)	0.67	<0.001
EDSS – Sensory Functional System *	0 (0 to 5)	0.45	0.001
10 metre walk (seconds) ** #	13 (14)	0.87	<0.001
Disability Measures			
*FIM – Transfer (Items I to K)**	20 (12 to 21)	-0.72	<0.001
FIM – Locomotion (Items L and M) *	12 (2 to 14)	-0.73	<0.001
Ambulation Index	2 (0 to 9)	0.90	<0.001
Perceived lower limb disability (VAS) **	65 (28)	-0.80	<0.001
Handicap measures			
London Handicap Scale – Mobility **	3 (2.7)	-0.43	0.002
Health Related Quality of Life			
SF 36 – Physical Functioning **	20 (6)	-0.87	<0.001

* Median (range); ** Mean (SD); # Number = 44 patients

As the strongest correlations were established with other disability rather than impairment, handicap, or heath related quality of life measures, these results support the validity of the lower limb disability sub-scale as a measure disability.

H). Bladder disability sub-scale

The GNDS-R bladder disability sub-scale correlated moderately with the special category items of the SNRS (r = -0.76), and highly with the bladder and bowel EDSS Functional System (r = 0.82). It also correlated moderately with the bladder management item of the FIM (r = -0.70), and with the perceived degree of bladder disability (r = -0.73). No handicap or quality of life measures were available for comparison (Table 6.35).

Table 6.35 The correlation between the GNDS-R bladder disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Bladder disability sub-scale *	2 (0 to 4)	-	-
Impairment measures			
SNRS – Special category *	-3 (-7 to 0)	-0.76	<0.001
EDSS – Bladder and Bowel Functional System *	1 (0 to 4)	0.82	<0.001
Disability Measures			
FIM – Bladder management *	6 (1 to 7)	-0.70	<0.001
Perceived disability (VAS) **	73 (27)	-0.73	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (SD)

The high correlation between this sub-scale and the bladder and bowel EDSS Functional System and the special category of the SNRS is likely to reflect the relatively mild degree of bladder disability in this cohort and the nature of this sub-scale which concentrates on the symptomatic aspects of bladder dysfunction in its lower grades. However, its high correlation with other disability measures lends support to its validity as a disability measure.

I). Bowel disability sub-scale

The GNDS-R bowel disability sub-scale correlated moderately with the special category items of the SNRS (r = -0.53), and the bladder and bowel EDSS Functional System (r = 0.65), and highly with bowel management item of the FIM (r = -0.84), and the perceived degree of bowel disability (r = -0.78). No handicap or quality of life measures were available for comparison (Table 6.36).

Table 6.36 The correlation between the GNDS-R bowel disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***P***
Bowel disability sub-scale *	0 (0-5)	-	-
Impairment measures			
SNRS – Special category *	-3 (-7 to 0)	-0.53	<0.001
EDSS – Bladder and Bowel Functional System *	1 (0 to 4)	0.65	<0.001
Disability Measures			
FIM – Bowel management *	7 (6 to 7)	-0.84	<0.001
Perceived bowel disability (VAS) **	84 (23)	-0.78	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (S D)

As the strongest correlation was established with disability rather then impairment measures, these results support the validity of this sub-scale as a measure of disability.

J). Fatigue disability sub-scale

The GNDS-R fatigue disability sub-scale correlated moderately with the overall score of the Chalder fatigue scale (r = 0.57), and with its physical fatigue domain (r = 0.54), but weakly with its mental fatigue domain (r = 0.36). It also correlated moderately with the perceived degree of fatigue disability (r = -0.70),

but weakly with the vitality domain of the SF-36 (r = -0.46). No handicap measures were available for comparison (Table 6.37).

Table 6.37 The correlation between the GNDS-R fatigue disability sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***r***	***p***
Fatigue disability sub-scale *	2 (0 to 4)	-	-
Impairment measures			
Chalder Fatigue Scale *			
Total score	0 (0 to 13)	0.57	<0.001
Physical fatigue	0 (0 to 8)	0.54	<0.001
Mental fatigue	0 (0 to 6)	0.36	0.029
Disability Measures			
Perceived fatigue disability (VAS) **	71 (25)	-0.70	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
SF 36 Vitality **	42 (21)	-0.46	0.001

* Median (range); ** Mean (SD)

As the strongest correlation was established with disability rather than impairment or health related quality of life measures, these results support the validity of the GNDS fatigue disability sub-scale as a disability measure, and indicate that it is tapping physical rather than mental fatigue.

K). Sexual disability sub-scale

1). Male patients

Only 11 male patient (57.9% of the total male population) consented to completing the GRISS questionnaire. Other patients felt that the questionnaire was too detailed and too personal to be completed. The median (range) of the sexual disability sub-scale scores and of the GNDS-R sum scores of the respondents and the non-respondents were identical at 3 (0 to 5) and 17 (0 to 28) respectively. The GNDS-R sexual disability sub-scale correlated moderately with the special category items of the SNRS (r = -0.54), and highly with the GRISS

questionnaire overall score (r = 0.85), and its impotence (r = 0.89) and non-communication (r = 0.87) domains, and moderately with its premature ejaculation (r = 0.63), avoidance (r = 0.62), dissatisfaction (r = 0.72) and infrequency (r = 0.61) domains. It also correlated moderately with the perceived degree of sexual function disability (r = -0.62). No other handicap or quality of life measures were available for comparison (Table 6.38).

Table 6.38 The correlation between the GNDS-R sexual disability sub-scale of male patients and other disablement and health related quality of life measures

	Score	***r***	***p***
Sexual disability sub-scale * #	3 (0 to 5)	-	-
Impairment measures			
SNRS – Special category * #	-3 (-7 to 0)	-0.54	0.018
Disability Measures			
GRISS questionnaire * # #			
Overall score	5 (1 to 9)	0.85	0.004
Impotence	6 (2 to 9)	0.89	0.004
Premature ejaculation	6 (1 to 9)	0.63	0.070
Non-sensuality	3 (1 to 7)	0.39	NS
Avoidance	1 (1 to 7)	0.62	0.078
Dissatisfaction	2 (1 to 7)	0.72	0.001
Infrequency	6 (1 to 9)	0.61	0.083
Non-communication	4 (1 to 7)	0.87	0.003
Perceived sexual disability (VAS) ** #	68.1 (29.3)	-0.62	0.007
Handicap measures			
NA			
Health Related Quality of Life			
NA			

* Median (range); ** Mean (SD); # n = 19 patients; ## n = 11 patients

As the strongest correlations were established with other disability rather than impairment measures, these results support the validity of this sub-scale as a measure of disability.

2). Female patients

Only 19 female patient (61% of the total female population) completed the GRISS questionnaire. Other patients felt that the questionnaire was too detailed and too personal to be completed. The median (range) of the sexual disability sub-scale scores and of the GNDS-R sum scores of the respondents and the non-respondents were identical at 0 (0 to 4) and 9 (0 to 26) respectively. The GNDS-R sexual disability sub-scale correlated moderately with GRISS questionnaire infrequency ($r = 0.56$) and anorgasmia ($r = 0.59$) domains, and with the perceived degree of sexual function disability ($r = -0.51$), but weakly with the special category items of the SNRS ($r = -0.42$), and with the GRISS questionnaire overall score ($r = 0.37$) and its vaginismus domain ($r = 0.35$). No other handicap or quality of life measures were available for comparison (Table 6.39).

Table 6.39 The correlation between the GNDS-R sexual disability sub-scale of female patients and other disablement and health related quality of life measures

	Score	***r***	***p***
Sexual disability sub-scale * #	0 (0 to 4)	-	-
Impairment measures			
*SNRS – Special category * #*	-3 (-7 to 0)	-0.42	0.022
Disability Measures			
*GRISS questionnaire * # #*			
Overall score	3 (1 to 3)	0.37	0.040
Infrequency	5 (3 to 9)	0.56	0.024
Non-communication	4 (1 to 9)	0.06	NS
Dissatisfaction	3 (1 to 4)	0.16	NS
Avoidance	2 (1 to 6)	0.35	NS
Non-sensuality	3 (1 to 7)	0.22	NS
Vaginismus	1 (1 to 6)	0.35	0.040
Anorgasmia	7 (2 to 9)	0.59	0.017
*Perceived sexual disability (VAS) ** #*	89.2 (17.4)	-0.51	0.011
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
NA	-	-	-

* Median (range); ** Mean (SD); # $n = 31$ patients; ## $n = 19$ patients

Although female patients recruited for this study had relatively low sexual disability grades in comparison to male patients, these results support the validity of this sub-scale as a measure of sexual disability as the strongest correlation was established with other measures of disability rather than impairment, handicap, or quality of life.

L). Other disabilities sub-scales

The GNDS-R other disability sub-scale correlated moderately with the perceived degree of other disabilities (r = -0.72), but only weakly with the pain item of the SF 36 (r = -0.35). No other impairment or handicap measures were available for comparison (Table 6.40).

Table 6.40 The correlation between the GNDS-R other disabilities sub-scale and other disablement and health related quality of life measures (n = 50)

	Score	***p***	***r***
Other disabilities sub-scale *	1 (0 to 4)	-	-
Impairment measures			
NA	-	-	-
Disability Measures			
*Perceived other disability (VAS)***	83.5 (24.1)	-0.76	<0.001
Handicap measures			
NA	-	-	-
Health Related Quality of Life			
SF 36 – Bodily pain **	76 (25)	0.35	0.014

* Median (range); ** Mean (SD)

As the strongest correlation was established with disability rather than health related quality of life measures, these results support the validity of this sub-scale as a disability measure.

M). GNDS-R sum score

The GNDS-R sum score correlated moderately with the SNRS (r = -0.75) and the EDSS (r = 0.78) as impairment measures. The GNDS-R also correlated highly with the FIM (r = -0.84), and moderately with the disability domain of the CAMBS (r = 0.73), the Barthel Index (r = -0.76), Ambulation Index (r = 0.74), and patients' perceived degree of disability (r = 0.75) as disability measures, and moderately with handicap domain of the CAMBS (r = 0.65) and with the London Handicap Scale (r = -0.52) as handicap measures. The GNDS-R sum score also correlated highly with the physical functioning domain of the SF-36 (r = -0.81), and moderately with the physical role limitation (r = -0.57) and the vitality domains (r = -0.58) of the SF-36 and with the overall quality of life as assessed by the EuroQol (r = -0.61), but weakly with the other domains of the SF-36 as health-related Quality of life measures (Table 6.41).

As the strongest correlations were established with other disability rather than impairment, handicap, or heath related quality of life measures, these results support the validity of the GNDS-R as a measure of disability.

6.8.3 Group differences and hypothesis testing

The two disability rank lists, which were compiled by myself and the research nurse, were almost identical (r = 0.99, p = <0.001). The GNDS-R sum score correlated highly with the mean ranks of disability (r = 0.91), moderately with the patients' ability to work (r = 0.71) and do their house work (r = 0.59), and weakly with the degree of patient's independence (r = 0.43) (Table 6. 42).

Thirty-seven patients had close relatives or carers able to complete the GNDS-R within three days of patients' clinical assessments. The correlation between the GNDS-R scores obtained by interview scale administration and the scores obtained when the scale was completed by patients' close relatives or carers to reflect their observed or perceived degree of disability independently of patients' own assessments of their disability was very high for the sum score and the majority of its sub-scales (r = 0.82 to 0.94) but moderate for the mood (r = 0.77), fatigue (r = 0.77), and other disabilities (r = 0.72) sub-scales (Table 6.43).

Table 6.41 The correlation between the GNDS-R sum score and other disablement and health related quality of life measures

	Score	***r***	***p***
GNDS sum score *	12 (1 to 29)	-	-
Impairment measures			
SNRS *	69 (25 to 98)	-0.75	<0.001
EDSS *	4.5 (0 to 7.5)	0.78	<0.001
Disability measures			
FIM *	121 (91 to 126)	-0.84	<0.001
Barthel *	20 (9 to 20)	-0.76	<0.001
CAMBS – disability *	2 (1 to 4)	0.73	<0.001
Ambulation Index *	2 (0 to 9)	0.74	<0.001
Perceived disability (VAS) **	75 (21)	0.75	<0.001
Handicap measures			
CAMBS – Handicap *	2 (1 to 4)	0.65	<0.001
London Handicap Scale **	60 (16)	-0.52	0.001
Quality of Life measures			
EuroQol VAS **	72 (22)	-0.61	<0.001
SF 36: **			
Physical functioning	20 (6)	-0.81	<0.001
Physical role limitation	50 (42)	-0.57	<0.001
Emotional role limitation	69 (42)	-0.35	0.02
Social functioning	62 (25)	-0.49	<0.001
Mental health	72 (19)	-0.39	0.007
Vitality	42 (22)	-0.58	<0.001
Bodily pain	76 (25)	-0.43	0.002
General health perception	48 (25)	-0.47	0.001

* Median (range); ** Mean (SD)

Table 6.42 GNDS-R group difference and hypothesis testing construct validity

	r	*p*
Work	0.71	<0.001
House work	0.59	<0.001
Independence	0.43	0.001
Disability rank	0.91	<0.001

Although patients' close relatives or carers overestimated overall, fatigue, and other disabilities and underestimated lower limb disability compared with the GNDS-R disability scores, these results suggest that the interview administration of the GNDS-R, which is based on patients' reports, is a valid method of applying this scale as it is capable of reflecting the degree of patients' disability as perceived by their close relatives or carers. These results therefore support the objectivity of the GNDS-R as a disability measure.

Table 6.43 The correlation between close relative or carers independent disability assessment and disability scores obtained through interview administration of the GNDS-R ($n = 37$)

Scale items	Interview median (range)	Relative /carer median (range)	*r*
Cognitive disability	0 [0 to 4]	0 [0 to 5]	0.83
Mood disability	0 [0 to 3]	0 [0 to 3]	0.77
Visual disability	0 [0 to 3]	0 [0 to 3]	0.83
Speech disability	0 [0 to 3]	0 [0 to 3]	0.82
Swallowing disability	0 [0 to 2]	0 [0 to 2]	0.94
Upper limb disability	1 [0 to 4]	1 [0 to 4]	0.93
Lower limb disability	3 [0 to 5]	2 [0 to 5]	0.92
Bladder disability	0 [0 to 5]	0 [0 to 5]	0.83
Bowel disability	0 [0 to 4]	0 [0 to 5]	0.89
Fatigue	2 [0 to 4]	3 [0 to 4]	0.77
Sexual disability	0 [0 to 5]	0 [0 to 5]	0.89
Other disabilities	0 [0 to 4]	1 [0 to 3]	0.72
GNDS-R sum score	12 [0 to 28]	15 [0 to 29]	0.92

6.9 The validity of the GNDS-R sum score

A sum score is a desirable attribute for any outcome measure. It simplifies statistical analysis and allows direct comparison between different patients. The validity of summing the GNDS-R sub-scales to obtain a sum score and the need for a weighting system was therefore assessed using two methods: regression analysis and factor analysis.

6.9.1 Regression analysis

In multiple regression, the values of one variable (the dependent variable: y) are estimated from those of two or more other variables (the independent variables: x_1, x_2, …., x_p) (Norman and Streiner, 1993c). This is achieved by the construction of a linear equation (the multiple linear regression equation) of the general form:

$$Y = b_1 (x_1) + b_2 (x_2) + ... + b_p (x_p) + b_0$$

where the parameters b_1, b_2, .., b_p are the partial regression coefficients and the intercept b_0 is the regression constant.

Using the inter-rater reliability data-set, the 12 GNDS-R sub-scales were regressed against the patients' perception of the degree of their disability as assessed by the 100-point visual analogue scales adopted from the EuroQol, and the patients' disability ranks which were compiled by myself as described in the validity study. The analysis suggested variable beta values for the different sub-scales (Table 6.44) which were used to obtain weighted sum scores using the multiple regression equation. The correlation between the raw and the weighted scores were high ($r = 0.95$ to 0.97, $p = <0.001$) suggesting that a weighting system was not needed.

Table 6.44 GNDS-R regression analysis

	Disability ranks (doctor)	Perception of disability (patient)
Regression sum square	16235.90	16681.54
Residuals sum square	2332.09	5110.29
Multiple regression coefficient	0.95	0.88
Regression constant	98.57	98.36
Independent variables		
Beta- Cognitive disability	0.04	0.04
Beta-Mood disability	0.08	0.22
Beta-Visual disability	0.23	0.28
Beta-Speech disability	0.11	0.37
Beta-Swallowing disability	0.22	0.34
Beta-Upper limb disability	0.18	0.05
Beta-Lower limb disability	0.48	0.29
Beta-Bladder disability	0.10	0.06
Beta-Bowel disability	0.14	0.08
Beta-Fatigue	0.07	0.22
Beta-Sexual disability	0.13	0.04
Beta-Other disabilities	0.01	0.05
Correlation between raw and weighted scores	0.96	0.97

6.9.2 Factor analysis

Vickrey and co-workers (1993) suggested a statistical method for reaching an overall score in multidimensional scales, which takes into account the possibility that different domains may contribute unequally to the construct being assessed. The suggested method was applied to the inter-rater reliability data set as follows:

1). Based on the oblique four factor rotated solution for the GNDS-R, four composite scores were created by weighting and summing the appropriate sub-scores. The specific weights contributing to each composite score were derived from dividing each sub-scale's factor loading by the sum of the factor loading of all the sub-scales selected as contributing to that factor's composite score.

2). These four composite scores were averaged to obtain a summary score.

3). The summary score was regressed onto the 12 GNDS-R sub-scores, and the standardised beta coefficients of this regression analysis were adjusted for a

theoretical maximum GNDS-R score of 100, and were used to assign relative weights to each GNDS sub-scale.

4). The sum of each sub-scale score times its weight yielded a weighted GNDS-R sum score (Table 6.45).

Table 6.45 The GNDS-R Vickrey weighting system

Composite	Formula (Using the 12 GNDS-R sub-scales and the suggested weights)	
Spinal disability	= 1.5	Lower limb disability score
	+ 1.8	Bladder disability score
	+ 1.8	Bowel disability score
	+ 1.5	Sexual disability score
Mental disability	= 2.2	Cognitive disability score
	+ 2.1	Mood disability score
	+ 2.2	Fatigue disability score
Bulbar disability	= 1.4	Speech disability score
	+1.3	Swallowing disability score
Arm/vision/other disabilities	= 1.6	Upper limb disability score
	+ 1.0	Vision disability score
	+ 1.6	Other disabilities score
Overall score	=	Sum of all 12 weighted sub-scales (0 – 100 point scale)

The correlation between the raw and the weighted sum scores was very high ($r = 0.99$, $p = <0.001$) suggesting that this weighting system has complicated the calculation process but not added to the validity of the final score.

6.10 Discussion

Multiple sclerosis is a multidimensional disease characterised by a wide variability of clinical manifestations and natural history. Clinical outcome measures used in this illness should therefore have relevant scale items, be able to embrace all its clinical manifestations, and have high levels of reliability, validity, and responsiveness. The GNDS was devised as a comprehensive clinical disability scale capable of fulfilling the need for a new outcome measure which is meaningful, relevant to patients' disability experiences, practical to administer, cost effective, and psychometrically sound.

At a conceptual level, the GNDS was devised as a disability scale, to complement rather than replace other impairment, handicap, and health-related

quality of life scales. Disability was thought to be particularly important due to its direct and practical relevance to the patients' ability to perform their various activities of daily living, and its indirect repercussion on health care resources and society at large.

A detailed review of the literature, supplemented by open interviews with 5 patients with multiple sclerosis and advice from a panel of experts, suggested that disability in multiple sclerosis is a multidimensional construct that can be categorised in 12 mutually exclusive dimensions which included cognition, mood, vision, speech, swallowing, upper limb, lower limb, bladder, bowel, fatigue, sexual function, and other disabilities to include pain, vertigo, and spasms. The severity of disability in each of these 12 dimensions was graded according to its impact on this particular function and the need for assistance according to 7-point severity scale. This process resulted in a comprehensive 12-category disability scale, capable of embracing the wide range of possible disabilities that could be experienced by patients with multiple sclerosis through the various stages of its natural history. To facilitate the application of the GNDS and to improve its reproducibility, an additional set of 12 interview sections was devised to complement the scoring sections. The interview sections contained sets of standard questions designed to ascertain the presence and the severity of disability according to the relevant disability sub-scales. This process created a 12- item comprehensive disability scale which could be administered by patient interview.

The face and content validity of this pilot scale were assessed by asking a cohort of 49 international experts to review the scale and indicate the degree of their approval or disapproval using a standard questionnaire. The majority of the referees approved the scale confirming its face validity. Their critical comments in relation to the scale contents were utilised to modify the scale as discussed later. The GNDS was piloted on a cohort of 64 patients with a wide rage of disabilities to assess its inter-rater reliability and construct validity, and a sub-group of 50 were followed up for 6 months to assess its intra-rater reliability and responsiveness. This study showed the GNDS to be internally consistent, have high inter- and intra-rater reliability, and to be moderately responsive to clinical change. However the frequency distribution of the scale was skewed to the 'less disabled' end of the scale suggesting a 'floor' effect.

This pilot scale therefore needed to be revised, and this process was achieved by utilising the critical comments of the referees which were collected in the face and content validity study. The revisions included reducing the number of the grades in each sub-scale from 7 to 6 (between 0 and 5) and modifying the disability grades of all 12 sub-scales to create the Revised Guy's Neurological Disability Scale (GNDS-R). Such a major revision meant that the psychometric data already collected were no longer applicable and that the scale's psychometric properties needed to be evaluated again. Ideally, such an evaluation needed to be done on a naive cohort of patients to avoid any bias resulting from training effects. However, due to administrative difficulties in recruiting and following a further cohort of patients, the second evaluation was done on the group of 50 patients with mild to moderate disability who had participated in the intra-rater reliability and responsiveness assessment of the GNDS. These patients were followed up for 9 months with three monthly assessments. Such a strategy was felt to be acceptable because the revisions were of such magnitude as to have effectively resulted in a new scale. Furthermore the previous administrations of the GNDS had been separated by three monthly intervals which minimised any subject or rater bias resulting from the effect of training or recall of previous answers. My previous study of the psychometric properties of other clinical scales used for multiple sclerosis suggested that raters' and patients' bias was insignificant when these scales were applied at three monthly intervals (Sharrack et al., 1999c).

The face validity of the revised scale was reassessed by inviting the same group of 49 international experts who took part in the first face and content study to examine the GNDS-R critically and indicate the degree of their approval or disapproval of the scale and its 12 categories. Compared with the first face validity study, more referees approved the revised scale. Eighty two percent of the respondents expressed their approval of the scale in general, and the majority also approved the various sub-scales therefore supporting the validity of this scale as a disability measure for multiple sclerosis.

Internal consistency of the GNDS-R was high with a Cronbach's alpha of 0.79 indication that the 12 sub-scales were internally consistent. Factor analysis suggested four meaningful factors which accounted for 58.7% of the total variance. The rotated matrix suggested no redundant items as each sub-scale loaded on one factor only suggesting that that the 12 GNDS categories are tapping

separate and mutually exclusive domains of human function. The first factor of the rotated matrix (*spinal factor)* correlated with the lower limb, bladder, bowel, and sexual function sub-scales. The second factor (*mental factor)* correlated with the cognition, mood, and fatigue sub-scales. The third factor (*bulbar factor*) correlated with the speech and swallowing sub-scales. The fourth factor (*upper limb / vision / other disabilities factor*) correlated with the upper limb, vision, and other disabilities sub-scales.

Inter-rater reliability of the different GNDS-R sub-scales was high with kappa coefficients ranging between 0.54 and 1 (moderate to perfect), and intraclass correlations coefficients ranging between 0.82 and 1 (almost perfect to perfect). Intra-rater reliability was also high with kappa coefficients ranging between 0.46 and 0.87 (moderate to almost perfect), and intraclass correlation coefficients ranging between 0.77 and 0.96 (substantial to almost perfect). Reliability of the GNDS-R sum scores was also very high with inter- and intra-rater intraclass correlation coefficients of 0.98 and 0.96 (almost perfect) respectively. Complete inter- and intra-rater agreement were obtained by allowing a difference of 3 and 5 points respectively.

The GNDS-R sum score was sensitive to clinical change with an effect size of 0.58 (moderate). In addition, the mood, vision, upper limb, lower limb, bladder, fatigue, and other disabilities sub-scales were also responsive with effect size values ranging between 0.23 (small) to 0.92 (large). The other sub-scales, cognition, speech, swallowing, bowel, and sexual, were unresponsive in this cohort reflecting the relatively static nature of these disabilities.

The frequency distribution of the GNDS-R sum scores in the studied cohort was slightly skewed to the normal end of the scale reflecting the range of disability (mild to moderate, EDSS: 0 to 7.5) of the patients in this study. The frequency distribution therefore needed to be reassessed in a more representative cohort.

The GNDS-R was designed as a simple instrument which could be applied by any health care personnel during an interview, over the telephone, or via a postal questionnaire so as to simplify the conduct and reduce the cost of clinical trials. Inter-rater reliability of the scale when applied by a neurologist and a nurse, or by a neurologist and a patient's relative or carer was therefore tested and found to be very high. The GNDS-R sum score interclass correlation coefficients

of the neurologist – nurse scale administration was 0.96 (almost perfect) with sub-scales kappa coefficients ranging between 0.58 and 0.95 (moderate to almost perfect). The GNDS-R sum-score intraclass correlation coefficient of the neurologist – relative scale administration was equally high at 0.91 (almost perfect) with sub-scales kappa coefficients ranging between 0.24-0.72 (fair to substantial). The GNDS-R was also found to be valid and reliable when administered over the telephone or via a postal questionnaire. With the exception of the mood disability sub-scale, the correlations between the scores obtained by administering the scale during an interview or over the telephone by the same rater were high with Spearman rank correlation coefficients ranging between 0.84 and 1. The correlations between the scores obtained by administering the scale during an interview or via a postal questionnaire were equally high for the GNDS-R sum scores and the majority of the sub-scales (speech, swallowing, upper limb, lower limb, bladder, bowel, fatigue, and sexual function) with Spearman rank correlation coefficients ranging between 0.81 and 1, but moderate for the cognition, mood, visual, and other disabilities sub-scales with Spearman Rank correlation coefficients ranging between 0.66 and 0.76. When compared with the interview administration, the GNDS-R telephone scale administration was very reliable with sum score interclass correlation coefficients of 0.96 (almost perfect) and sub-scales kappa coefficients ranging between 0.69 and 1 (substantial to almost perfect). GNDS-R sum score intraclass correlation coefficient of the postal questionnaire administration was equally high at 0.93 (almost perfect) with sub-scales kappa coefficients ranging between 0.49 and 0.84 (substantial to almost perfect).

Records of the time needed to administer the GNDS-R by various raters (neurologist, nurse, patient, patient's relative) and by various methods (interview, over the telephone, via a postal questionnaire) showed that the scale could be applied on average in <10 minutes suggesting that the scale constitute a minor burden for patients and raters.

The validity of summing the GNDS-R sub-scales to obtain a sum score and the need for a weighting system was assessed using regression analysis and factor analysis. Both methods showed high correlations between the raw and the weighted sum scores suggesting that a weighting system was an unnecessary complication.

The postal version of the GNDS-R was later tested on 194 naive patients to evaluate the performance of this scale in such large community based cohorts and to assess its acceptability to patients. Eighty three percent of the returned questionnaires were fully and correctly completed and an additional 15% contained enough data to allow assigning sum scores, suggesting that the scale's layout was simple, user friendly and acceptable to patients. The frequency distribution of the GNDS-R sum scores was Normal with no ceiling or floor effects. Eighty eight percent of the respondents expressed their approval of the scale in general and the majority expressed their approval of the different sub-scales supporting the face validity of this scale from the patients' perspective.

The face validity of the GNDS-R as a measure of disability was confirmed by demonstrating its high correlation with another disability measures including the FIM, patients disability ranks, and patients' self-assessment of disability using the physical functioning domain of the SF-36, and its moderate correlation with the EDSS, the disability domain of the CAMBS, the Ambulation Index, and the degree of disability as perceived by the patients. The moderate correlation between the GNDS-R and the Barthel Index is likely to be due to generic differences between these two scales. The Barthel Index is not a disease specific scale. It addresses disability in a limited range of activities of daily living (Table 5.3), and is known to have 'floor' and 'ceiling' effects (Applegate et al., 1990; Wade, 1995b). The GNDS-R, on the other hand, is a comprehensive disease specific scale with no 'floor' or 'ceiling' effects as demonstrated in the postal survey study. As expected in any disability scale, the GNDS-R correlated moderately with other measures of impairment (SNRS), handicap (the London Handicap Scale) and quality of life (the EuroQol and the SF-36), and with patients' ability to work and do their housework. Interestingly, the correlation between the GNDS-R sum score and the handicap domain of the CAMBS was of a magnitude similar to its correlation with the EuroQol rather tan the London Handicap Scale suggesting that the handicap domain of the CAMBS is tapping health related quality of life rather than handicap.

The validity of the GNDS-R interview format, which depends on patients' level of reporting, was supported by the high correlation between the scores obtained during interview administration and patients' relatives or carers' independent perception of the degree of their disabilities. This correlation was

very high for the GNDS-R sum score ($r = 0.92$) and the cognitive, visual, speech, swallowing, upper limb, lower limb, bladder, bowel, and sexual disability sub-scales ($r = 0.82$ to 0.94) and moderate for the mood, fatigue, and other disabilities sub-scales ($r = 0.72$ to 0.77). The 'objectivity' of the GNDS-R is not surprising because its self-reporting elements are not based on any value judgements (as in the case of health related quality of life measures) but on directly observable or easily ascertainable dimensions which can be verified against a detailed history obtained from the patients, their families, close friends or carers.

The validity of the various GNDS-R sub-scales was also established by demonstrating high correlations between these sub-scales and other relevant measures of disability, and moderate correlations between them and other relevant measures of impairment, handicap, and health related quality of life.

The cognitive sub-scale correlated highly with the memory domain of the FIM and moderately with the perceived degree of cognitive disability. Rudick and co-workers advocated the use the 3 or the 2 second version of the PASAT as part of a composite outcome measure (Rudick et al., 1996b). In my study, the 4 second (but not the 2 second) version of the PASAT correlated weakly with the GNDS-R cognitive sub-scale confirming my earlier impression (chapter 3) that such composite outcomes are likely to be of limited value as they comprise impairment measures which do not adequately reflect patients' functional status.

The highest correlation in the mood disability sub-scale was demonstrated with the perceived degree of mood disability. Weaker correlations were established between this sub-scale and other impairment (including the General Health Questionnaire) and health related quality of life measures.

The highest correlation in the visual disability sub-scale was demonstrated with the perceived degree of visual disability. Weaker correlations were established between this sub-scale and other impairment measures including visual acuity, and the relevant items of the SNRS and the visual Functional System of the EDSS.

The speech disability sub-scale correlated moderately with the perceived degree of speech disability, and the lower cranial nerves domain of the SNRS, but weakly with other impairment measures including the Frenchay Dysarthria Assessment. The high correlation between this sub-scale and the lower cranial nerves domain of the SNRS is likely to reflect the mild degree of speech disability

in the cohort studied as this sub-scale concentrates on the symptomatic aspects of the speech disturbance in its less severe grades.

The highest correlation in the swallowing disability sub-scale was demonstrated with the perceived degree of speech disability. Weaker correlations were established between this sub-scale and other impairment measures including the timed swallowing test and the relevant items of the SNRS and the EDSS brain stem Functional System.

The upper limb disability sub-scale correlated moderately with the perceived degree of upper limb disability and the self-care items of the FIM. Weaker correlations were also established between this sub-scale and other impairment measures including the 9-peg hole test and the relevant items of the SNRS and the Functional Systems of the EDSS. Three patients were unable to perform the 9-peg hole test, the second component of the composite outcome measure advocated by Rudick and co-workers, suggesting a 'ceiling' effect and throwing doubt on its usefulness in clinical trials.

The lower limb disability sub-scale correlated highly with the Ambulation Index, patients' self-assessment of disability using the physical functioning domain of the SF-36, and the perceived degree of lower limb disability. It also correlated highly with the 10 metre walk, the relevant items of the SNRS, and the pyramidal Function System of the EDSS, and moderately with the transfer and locomotion items of the FIM. Six patients were unable to perform the 10 metre walk, which is very similar to the 25-foot walk suggested by Rudick and co-workers as the third component of the composite outcome measure, suggesting a 'ceiling' effect and rendering this measure inappropriate for clinical trials of multiple sclerosis.

The bladder disability sub-scale correlated moderately with the special category item of the SNRS, and highly with the EDSS bladder and bowel Functional System, the bladder management item of the FIM, and the perceived degree of bladder disability. The high correlation between this sub-scale and the relevant EDSS and SNRS items is likely to reflect the nature of this sub-scale which concentrates on the symptomatic aspects of bladder dysfunction as well as its consequences suggesting that it may be assessing both impairment and disability. However, its high correlation with other disability measures lends strong support to its validity as a measure of disability.

The bowel disability sub-scale correlated highly with the bowel management item of the FIM and moderately with the perceived degree of bowel disability. Weaker correlations were also noted with the SNRS and the EDSS related items.

The highest correlation in the fatigue disability sub-scale was demonstrated with the perceived degree of fatigue disability. Weaker correlation was also established between this sub-scale and Chalder Fatigue scale. Factor analysis showed this sub-scale to have loaded on the second factor along with the cognitive and the mood sub-scales, however its correlation with the physical domain of the Chalder Fatigue scale was higher than its correlation with the mental domain suggesting that its is tapping physical rather than mental fatigue.

Although 42% of male patients and 39% of female patients in this cohort did not complete the validity assessment of the sexual disability sub-scale, the median GNDS-R sum scores and sexual disability sub-scale scores of the participants and the non participants were identical indicating comparable degrees of disability. Among male patients, this sub-scale correlated highly with the GRISS questionnaire sum score and many of its items and moderately with the perceived degree of disability. However the correlation between this sub-scale and the GRISS questionnaire sum score and many of its items was weak for female patients, which is likely to be related to the previously reported low degree of sexual disability in this sub-group of patients (Valleroy and Kraft, 1984). Similar low correlation figures were obtained by using the self-administered version of the sub-scale (data from the postal questionnaire study) indicating that female patients did not underreport their sexual disabilities during the interview administration of the scale.

Finally the correlation between the other disability sub-scale and the perceived degree of disability was moderate supporting the validity of this sub-scale as a disability measure.

6.11 Raters' and patients' bias

As discussed in chapter 5, this study was designed to minimise the effect of raters' and patients' bias on the assessment of reliability and responsiveness. All the raters were blinded to their own and other raters' previous scores, and open discussions about patients' clinical conditions were avoided amongst

themselves. In the inter-rater reliability study, patients were assessed independently by the two raters and no fixed order for the examination was observed so as to reduce the effect of patients' bias which may result from practice effect or fatigue. Data for the intra-rater reliability and responsiveness study were collected at three monthly intervals so as to reduce raters' and patients' bias which may result from recall of the previous assessments. The effect of this potential source of bias is unlikely to have been significant since the inter-rater reliability figures were often higher than the intra-rater reliability figures. It is also unlikely that the familiarity of the patients to the assessors or the frequent administrations of the GNDS-R have biased the results since the reliability figures of the various methods of scale administration were often lower then the initial inter-rater reliability figures which were obtained when the scale was administered by the two raters for the first time. A noticeable deficiency in this study is the lack of supportive psychometric data by independent investigators on naive cohorts of patients. Such a study in now underway in six centres throughout the UK.

6.12 Conclusion

The Guy's Neurological Disability Scale goes a long way towards meeting the need for a new clinical disability scale for multiple sclerosis. It is a comprehensive multidimensional clinical rating scale capable of embracing the whole range of disabilities likely to be encountered in this illness. It is simple and user-friendly, acceptable to neurologists and patients, and can be applied by non-medically qualified health care staff or by the carers of patients without prior training. In our hands, this scale is reliable, responsive, and valid as a disability scale. As a clinical outcome measure for multiple sclerosis, this scale is capable of providing relevant information on the degree of disability as experienced by the affected patients. Additional assessment of impairment, handicap, and health related quality of life could be necessary to provide a more comprehensive appraisal of patients' health status.

Chapter 7

MAGNETIC RESONANCE IMAGING IN MULTIPLE SCLEROSIS

7.1 Introduction

The increasing use of Magnetic Resonance Imaging (MRI) over the last 10 to 15 years as a diagnostic and research tool has had a major impact on our understanding of the pathogenesis and the natural history of multiple sclerosis. MRI has greatly improved the diagnostic yields in patients with inconclusive clinical features, and has allowed the dynamic disease processes to be visualised as never before. The difficulty in developing satisfactory measures of clinical outcome in multiple sclerosis and the sensitivity of MRI in detecting disease activity compared with clinical relapse rate (Paty, 1988) has also led to its increasing use as a primary and secondary outcome measure in many phase II and III therapeutic clinical trials. However the lack of a consistent association between clinical outcome measures and conventional MRI parameters has been a major concern. T2-weighted brain imaging complimented with gadolinium enhanced T1-weighted imaging have traditionally been used for the diagnosis and the monitoring of disease activity, whereas unenhanced T1-weighted imaging has been of limited use.

7.1.1 T2 weighted imaging

Conventional T2-weighted spin echo pulse sequence has been used in most clinical trials of multiple sclerosis (Miller et al., 1998). On this sequence multiple sclerosis lesions demonstrate high intensity while background white matter appears generally dark. The introduction of fast spin echo imaging allowed the acquisition of qualitatively equivalent thin contiguous imaging sections in times similar to those used for thicker sections on conventional spin echo images (Thorpe et al., 1994). Earlier studies of post-mortem cadaver MRI scans and subsequent pathological examination have provided support for using T2-weighted imaging as a measure of the extent of demyelination (Paty and Moor,

1997). However the pathological specificity of T2-weighted signals is very low. Standard T2-weighted MRI simply reflects the extent of 'water protons' in the various tissues and all the pathological processes in multiple sclerosis lesions (oedema, inflammation, demyelination, axonal loss, and gliosis) are therefore represented as hyperintense areas on this sequence (McDonald et al., 1994).

7.1.2 T1-weighted imaging

T1-weighted spin echo imaging shows multiple sclerosis lesions as areas of low signal intensity, or black holes, in contrast to the background isointense white and grey matter. Acute black holes develop and resolve in parallel with onset and recovery from relapses (van Waesberghe et al., 1997), but some may persist and become chronic. Histopathological studies of chronic hypointense lesions revealed a strong correlation between the degree of hypointensity on post-mortem T1-weighted images and both axonal density (r = -0.72) and the degree of matrix destruction and widening of the extracellular space (r = 0.45) (van Walderveen et al., 1996). Additionally, changes on T2-weighted images in patients with secondary progressive multiple sclerosis were more frequently accompanied by changes on T1-weighted images compared with relapsing remitting patients suggesting that T1-weighted hypointense lesion load is the MRI equivalent of failure of remission (Truyen et al., 1996). Chronic hypointense T1-weighted lesions therefore represent the more disabling lesions compared with lesions which are only noted on T2-weighted images.

7.1.3 Other MRI parameters

The nature of the pathological process in multiple sclerosis can also be assessed using Gadolinium-enhanced T1-weighted imaging to examine the degree of inflammation and blood brain barrier breakdown (Stone et al., 1995), magnetization transfer imaging to assess the degree of demyelination and axonal degeneration (McGowan et al., 1997), diffusion weighted imaging to examine the integrity of myelin and fiber pathways (Horsfield et al., 1997), proton spectroscopy to assess the biochemical changes associated with myelin destruction (Arnold et al., 1997), and cerebral and spinal cord atrophy as a marker of tissue loss (Losseff et al., 1996a; Losseff et al., 1996b).

7.2 Correlation between MRI and clinical rating scales

Two main MRI methods have been used to evaluate disease activity in multiple sclerosis and to monitor treatment efficacy: counting the number of new or active brain lesions and measuring total brain lesion load.

Weak correlations (r = 0.19-0.23) have been reported between the number of new and active MRI lesions and the frequency of clinical relapses (Simon et al., 1998). The correlation between the presence of active lesions and long-term clinical evolution is less clear (Morrissey et al., 1993).

The presence and the extent of T2-weighted MRI abnormalities at first presentation with clinically isolated syndromes suggestive of multiple sclerosis correlate moderately with the degree of disability after 5 and 10 years (r = 0.75 and 0.45 respectively) (Morrissey et al., 1993; O'Riordan et al., 1997). However in established multiple sclerosis, the correlation between T2-weighted abnormalities and clinical disablement remains modest (Table 7.1). Cross-sectional studies have shown no (Thompson et al., 1990), or only modest correlations (r = 0.22-0.28) (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995) between T2-weighted lesion loads and clinical outcome measures of impairment and disability. However when the association between lesion load at specific anatomical sites and relevant clinical presentations is examined, higher correlations (r = 0.33-0.52) are found between cerebral and brain stem lesion load and the EDSS and its brain stem Functional System scores, the Ambulation Index, and other tests of upper extremity function (Baumhefner et al., 1990), and between cerebral lesion load and corpus callosum atrophy and cognitive dysfunction (Franklin et al., 1988; Rao et al., 1989). Many longitudinal studies have shown only modest correlation between the change in T2-weighted MRI lesion load and the sustained change in clinical disablement measures (r = 0.13-0.23) (The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, 1995; Filippi et al., 1995).

Table 7.1 Correlation between T2-weighted lesion load and clinical outcomes

Study	Correlation	*r (p)*
Gass (1994)	T2 lesion load and EDSS	0.33 (0.03)
IFNB MS Study Group (1995)	T2 lesion load and EDSS at entry	0.22 (<0.001)
	T2 lesion load and EDSS at exit	0.26 (<0.001)
	T2 lesion load and SNRS at entry	-0.25 (<0.001)
	T2 lesion load and SNRS at exit	-0.28 (<0.001)
	Change in T2 lesion load and change in EDSS	0.23 (<0.001)
	Change in T2 lesion load and change in SNRS	-0.21 (<0.001)
Van Walderveen (1995)	T2 lesion load and EDSS	0.30 (0.03)
	Change in T2 lesion load and change in EDSS	0.19 (NS)
Filippi (1995)	New T2 lesions and change in EDSS	0.13 (0.02)
	Enlarging T2 lesions and change in EDSS	0.18 (0.02)

Similar to T2-weighted lesion load, cross sectional analysis of T1-weighted lesion load showed no (Truyen et al., 1996) or only moderate correlation (r = 0.46) (van Walderveen et al., 1995) with the EDSS. However in longitudinal studies, the relative increase in T1-weighted lesion load correlated more strongly with disease progression as assessed by sustained change in the EDSS scores (r = 0.74-0.80) (van Walderveen et al., 1995; Truyen et al., 1996).

The use of new imaging techniques has also provided stronger correlations with clinical outcome measures. For example significant correlation (r = 0.70) has been found between EDSS and the diameter of the spinal cord at the level of C2 reflecting the importance of axonal damage on the progression of disability (Losseff et al., 1996b). Progressive cerebral atrophy has also been reported to correlate with increasing disability (Losseff et al., 1996a). Several other non-conventional MR putative markers of demyelination and axonal degeneration have been correlated more strongly with disability. Cross-sectional studies found moderate correlations between the degree of disability and average lesion Magnetisation Transfer ratio (r = -0.44) (Gass et al., 1994), and between decreased N-acetyle-aspartate resonance intensities in and around MS lesions and the degree of disability (r = -0.73) (De Stefano et al., 1995). The use of these non-conventional MR techniques in clinical trials requires further clinical validation in large longitudinal studies to confirm these encouraging initial results.

7.3 Quantitative MRI analysis techniques

Data obtained by quantitative MRI analysis are critical both for understanding the natural history of the disease and for monitoring the effects of the therapeutic interventions. The reliability of such data depends on the accuracy of two distinct processes: recognition (the identification of multiple sclerosis lesions), and delineation (defining the boundaries of such lesions). Human experts appear to be better than computers at recognition, whereas computers are better and faster at delineation (Miller et al., 1998). However both processes can be problematic. Multiple sclerosis lesions are often small in size and large in number, and have a 'lumpy-bumpy' appearance with irregular and fuzzy boundaries causing significant variation in their conspicuity and considerable difficulties in their delineation. There may also be many artefacts from scanners, patient motion, and other sources such as blood and CSF flow which could compromise the accuracy of lesion identification.

Many segmentation techniques have been devised and used in quantifying lesion load. Manual outlining is the current standard for lesion segmentation in almost all clinical trials (Miller et al., 1996). The accuracy of this technique is operator dependent with intra-rater variability as low as 6% and inter-rater variability as high as 14-40% (Paty and Li, 1993; Paty et al., 1994). Threshold-based segmentation methods substantially improve the reliability of volume measurement compared with manual tracing but they require considerable operator assistance to correct for false positives and negatives (Grimaud et al., 1996). Local (lesion-by-lesion) thresholding techniques, which use the advantage of expert recognition and computer delineation, offer a better precision but they take about as long as manual outlining to perform (Grimaud et al., 1996). Many other techniques with reduced human input have been developed in the past few years (Filippi et al., 1998). The main potential advantage of these techniques is to reduce the time and effort needed for the operator to assess the images and to improve reproducibility.

7.4 Rational for this study

The WHO ICIDH model predicts a close relation between pathology and impairment, but a weak relation between pathology and disability, and between

pathology and handicap (Wade, 1996). There are no published reports in the literature addressing the relation between pathology and the three ICIDH components in a single cohort of patients, or the relation between pathology and health related quality of life. This study was therefore designed to assess comprehensively the correlation between MRI lesion load in multiple sclerosis on two of the most commonly used acquisitions as a measure of pathology and various impairment, disability, handicap, and health relater quality of life outcome measures.

7.5 Patients and methods

Forty-eight patients with clinically or laboratory supported definite relapsing remitting or secondary progressive multiple sclerosis attending a multiple sclerosis outpatient research clinic were recruited for this study. The cohort consisted of 30 women and 18 men with a median age of 36 years (range 24-51), a median EDSS of 4.5 (range 0-7.5), and a median disease duration of 12 years (range 2-17). High resolution (slice thickness 2.4 mm, in plane resolution 0.89 by 0.89 mm) contiguous *T1*- (TR 28.3 ms, TE 6.9 ms) and *T2*-weighted (TR 3300 ms, TE 120 ms) MRI scans were performed on all patients using a standardised protocol on a 1.5 T Phillips ACS scanner at Guy's Hospital, London. All patients were assessed clinically on the same day as their scans and were assigned scores on the EDSS and the SNRS as impairment measures, the GNDS-R, the AI, the FIM, the disability components of the CAMBS and the Barthel Index as disability measures. Patients were also invited to complete the London Handicap Scale and the handicap components of the CAMBS as handicap measures, and the EuroQol and the Short Form 36 health survey questionnaire as health related quality of life measures within three days of their clinical assessment.

7.6 MRI analysis

Quantitative MRI analysis was done in a blinded fashion on a standard graphic workstation (Hewlett Backard 735), using a novel semi-automated integrated software package designed for the segmentation and analysis of multiple sclerosis lesions, Analysis Tool for MS (ATOMS), which was developed by Dr. Alan Colchester's research group at the United Medical and Dental

Schools, University of London. MRI images were stored on a local hard disc to allow rapid access while using ATOMS. A standard screen arrangement was employed in which three quarters were used to display slices of the image data while the fourth was used to select the method of segmentation, image display mode, image orientation (axial, coronal, sagital), zooming facility, slice numbering, and saving facilities. Alternative slice display could be selected to allow either a single image display or orthogonal display in the three cardinal anatomical planes of the anatomically corresponding levels. Several segmentation techniques are supported by ATOMS, including manual outlining, intensity based threshholding, and hierarchical segmentation. The latter method was used for this study. The hierarchical segmentation method is a novel clustering algorithm developed at the United Medical and Dental Schools by L.D. Griffin (Griffin et al., 1996) in which the user makes points inside and outside a lesion and the computer searches to find the largest object containing all the inside and none of the outside points. It works by grouping adjacent pixels which have low edge strength based on local grey level differences. Boundaries, which correspond to ridges of locally high edge strength, are established between pixels that belong to different clusters. The grouping is computed to produce a finely detailed segmentation. The hierarchical segmentation is fully automated and requires no thresholds or other adjustable parameters. Compared with manual editing and intensity based threshholding, the hierarchical segmentation method was found to shorten the time needed for interactive segmentation by at least 25% whilst providing comparable volumetric assessments (Colchester et al., 1996).

Five training sessions supervised by an experienced neuroradiologist (Dr. Tim Cox, Guy's Hospital) were arranged in which high-resolution T1- and T2-weighted scans of five patients with multiple sclerosis obtained using the same study protocol were segmented. All MRI analysis was done thereafter by myself in a blinded fashion. Multiple sclerosis lesions were identified and segmented using ATOMS, and each lesion was assigned one of three anatomical locations (cerebral, cerebellar, and brain stem) according to a reference atlas (Kretschmann and Weinrich, 1992). At the end of the analysis, six randomly selected T1- and T2-weighted scans were re-segmented (some three months after the initial segmentation) to assess intra-rater reliability. Inter-rater reliability was not assessed because this study involved a single rater.

7.7 Statistical analysis

Data were tabulated and analysed using SPSS 7.5 for windows. The EDSS, SNRS, FIM, AI, CAMBS, and Barthel Index were treated as ordinal data. The London Handicap scale, EuroQol, Short Form 36 health survey questionnaire and MRI lesion loads were treated as interval data. Intra-rater reliability was expressed as the mean and 95% confidence intervals of the difference between the two observations to assess rater bias, and the repeatability coefficient as an indication of the maximum difference required to achieve 95% rater agreement. Intra-rater reliability was also assessed using intra-rater variability calculated as the percentage difference between the first and the second observations divided by the first observation (van Walderveen et al., 1995). Correlation between lesion load and the various clinical scales was assessed using Pearson's and Spearman rank correlation coefficients for interval and ordinal data respectively.

7.8 Results

Depending on the extent of the MRI abnormalities, the analysis of each set of the T2-weighted images required on average 90 to 120 minutes, whereas the analysis of each set of the T1-weighted images required on average 60 to 90 minutes. Mean T2-weighted lesion loads were substantially higher than mean T1-weighted lesion loads in all three anatomical sites (Table 7.2). Ninety five percent of the total T2- and T1-weighted lesion loads was due to lesions in the cerebrum. Brain stem lesion load was higher than cerebellar lesion load on the T2-weighted images, whereas cerebellar lesion load was higher than brain stem lesion load on the T1-weighted images.

Table 7.2 Distribution of MRI lesion loads (cm^3)

Site	Mean	SD
T2-weighted imaging		
Total lesion load	14.37	18.66
Cerebral	13.59	18.31
Brain stem	0.67	0.96
Cerebellar	0.23	0.27
T1-weighted imaging		
Total lesion load	5.87	8.51
Cerebral	5.62	6.35
Brain stem	0.03	0.05
Cerebellar	0.16	0.21

7.8.1 Intra-rater reliability of lesion load analysis

Lesion loads in the six randomly selected scans analysed in this study were higher than mean lesion loads of the cohort as a whole (Table 7.3). Intra-rater reliability of the T2-weighted total, cerebral, brain stem, and cerebellar lesion loads was very high with intra-rater variability of 3.1% - 3.4%. Mean volume differences between the two assessments were small with narrow 95% confidence intervals which included the "0" value indicating the absence of rater bias. Repeatability coefficients of the total and cerebral lesion loads were small (11.8% and 13.2% of the first observation values respectively), but relatively high for the brain stem and cerebellar lesion loads (59.2% and 40.7% of the first observation values respectively).

By contrast, intra-rater reliability of the T1-weighted total and cerebral lesion loads was slightly lower with intra-rater variability of 14.5% and 14.1% respectively. Intra-rater reliability of the T1-weighted brain stem and cerebellar lesion loads was low with intra-rater variability of 52.5% and 35.6% respectively. Mean volume differences between the two assessments were slightly high with wide 95% confidence intervals suggesting a partial bias towards the second observation. Repeatability coefficients of the total and cerebral lesion loads were high (76.1% and 74.4% of the first observation values respectively), and very high for the brain stem and cerebellar lesion loads (122.2% and 137.5% of the first observation values respectively).

Table 7.3 Intra-rater reliability of lesion load analysis

	Time 1 Mean (SD) *	Time 2 Mean (SD) *	Mean (95% CI) difference *	Variability	RC *
T2-weighted					
Total lesion load	26.79 (42.09)	26.00 (40.44)	0.79 (-1.17 to 2.75)	3.1%	3.16
Cerebral	26.03 (42.04)	25.23 (40.41)	0.80 (-1.06 to 2.66)	3.4%	3.48
Brain stem	0.49 (0.23)	0.51 (0.31)	-0.02 (-0.18 to 0.14)	3.3%	0.29
Cerebellar	0.27 (0.32)	0.26 (0.33)	0.01 (-0.01 to 0.06)	3.1%	0.11
T1-weighted					
Total lesion load	10.81 (16.07)	12.38 (20.18)	-1.58 (-5.98 to 2.83)	14.5%	8.23
Cerebral	10.58 (15.83)	12.07 (19.77)	-1.49 (-5.70 to 2.73)	14.1%	7.87
Brain stem	0.09 (0.06)	0.14 (0.07)	-0.05 (-0.01 to 0.04)	52.5%	0.11
Cerebellar	0.32 (0.24)	0.44 (0.43)	-0.11 (-0.67 to 0.45)	35.6%	0.44

RC = repeatability coefficient; * cm^3

7.8.2 Correlation with clinical scales:

Patients' characteristics in terms of their impairment, disability, handicap, and health related quality of life scores are detailed in Tables 7.4 and 7.9.

A). Correlation with impairment outcome measures

Modest correlations were found between T2-weighted total lesion load and the cerebellar and brain stem Functional Systems; T2-weighted cerebral lesion load and the cerebellar and the brain stem Functional Systems; T2-weighted brain stem lesion load and the cerebellar and the brain stem Functional Systems; T2-weighted cerebellar lesion load and the cerebellar Functional System; and between T2-weighted cerebellar lesion load and the bladder and bowel Functional System (Table 7.4).

Table 7.4 Correlation between T2-lesion loads and the EDSS

Functional Systems	Score *	Correlation with MRI lesion loads			
		Cerebral	Brain Stem	Cerebellar	Total lesion load
Pyramidal	3 [0 to 5]	0.11	0.08	0.38	0.10
Cerebellar	2 [0 to 5]	0.29 (0.044)	0.41 (0.02)	0.59 (0.001)	0.30 (0.039)
Brain stem	0 [0 to 4]	0.42 (0.003)	0.56 (0.001)	0.29	0.48 (0.002)
Sensory	0 [0 to 5]	0.09	0.10	0.11	0.09
Bladder and bowel	1 [0 to 4]	0.11	0.04	0.49 (0.009)	0.11
Mental	0 [0 to 3]	0.27	0.24	0.14	0.27
Visual	0 [0 to 6]	0.22	0.18	0.11	0.23

* Median (range); figures in parenthesis represent the statistically significant *p* values

Modest correlations were also found between T1-weighted total lesion load and the cerebellar, brain stem, and mental Functional Systems; T1-weighted cerebral lesion load and the cerebellar, brain stem, and mental Functional Systems; T1-weighted brain stem lesion load and the brain stem Functional System; and between T1-weighted cerebellar lesion load and the cerebellar Functional System (Table 7.5).

Table 7.5 Correlation between T1-lesion loads and the EDSS

Functional Systems	Score *	Correlation with MRI lesion loads			
		Cerebral	Brain Stem	Cerebellar	Total lesion load
Pyramidal	3 [0 to 5]	0.18	0.16	0.39	0.18
Cerebellar	2 [0 to 5]	0.34 (0.019)	0.39 (0.032)	0.66 (0.022)	0.36 (0.013)
Brain stem	0 [0 to 4]	0.39 (0.006)	0.50 (0.012)	0.27	0.43 (0.003)
Sensory	0 [0 to 5]	0.15	0.22	0.11	0.16
Bladder and bowel	1 [0 to 4]	0.19	0.23	0.38	0.19
Mental	0 [0 to 3]	0.34 (0.017)	-0.13	0.09	0.35 (0.016)
Visual	0 [0 to 6]	0.29 (0.045)	-0.35	0.35	0.30 (0.042)

* Median (range); figures in parenthesis represent the statistically significant *p* values

Modest correlations were also found between T1- and T2-weighted total lesion loads and lesion loads in the different anatomical sites and the SNRS sum score (Tables 7.6 and 7.7).

Table 7.6 Correlation between T2-lesion loads and measures of impairment, disability, handicap, and health related quality of life

Scale item	Score	Correlation with MRI lesion loads			
		Cerebral	Brain Stem	Cerebellar	Total lesion load
Impairment measures					
SNRS	70 [36 to 98] *	-0.29 (0.039)	-0.40 (0.023)	-0.47 (0.013)	-0.30 (0.037)
EDSS	4.5 [0 to 7.5] *	0.17	0.24	0.41 (0.032)	0.16
Disability measures					
*GNDS**	12 [0 to 29] *	0.14	0.39 (0.027)	0.32	0.15
FIM	122 [103 to 126] *	-0.18	-0.32	-0.39 (0.039)	-0.19
Barthel	20 [9 to 20] *	-0.18	-0.04	-0.08	-0.17
CAMBS – disability	2 [1 to 4] *	-0.01	0.46 (0.008)	0.27	-0.01
Ambulation Index	2 [0 to 9] *	0.10	0.29	0.43 (0.025)	0.12
Handicap measures					
CAMBS – Handicap	2 [1 to 4] *	0.22	0.22	0.29	0.23
London Handicap Scale	61 [16] **	-0.08	0.09	-0.06	-0.08
Quality of Life measures					
EuroQol VAS	72 [22] **	-0.28	-0.28	-0.24	-0.28
SF 36:					
Physical functioning	20 [6] **	-0.11	-0.18	-0.36	-0.12
Physical role limitation	50 [42] **	-0.11	-0.17	-0.21	-0.09
Emotional role limitation	69 [42] **	-0.27	-0.08	-0.07	-0.26
Social functioning	62 [26] **	-0.18	-0.09	-0.14	-0.17
Mental health	72 [19] **	-0.19	-0.27	-0.29	-0.19
Vitality	42 [22] **	-0.07	-0.15	-0.15	-0.08
Bodily pain	76 [25] **	-0.00	-0.26	-0.08	-0.14
General health perception	48 [25] **	-0.16	-0.34	-0.25	-0.18

* Median (range); ** Mean (SD); figures in parenthesis represent the statistically significant *p* values

Table 7.7 Correlation between T1-lesion loads and measures of impairment, disability, handicap, and health related quality of life

Scale item	Score		Correlation with MRI lesion loads			
		Score	Cerebral	Brain Stem	Cerebellar	Total lesion load
Impairment measures						
SNRS	70 [36 to 98] *	70 (36 to 98) *	-0.36 (0.012)	-0.37 (0.029)	-0.46 (0.046)	-0.38 (0.008)
EDSS	4.5 [0 to 7.5] *	4.5 (0 to 7.5) *	0.23	-0.29	0.42 (0.034)	0.24
Disability measures						
*GNDS**	12 [0 to 29] *	12 (0 to 29) *	0.25	0.25 (0.038)	0.29	0.26
FIM	122 [103 to 126] *	122 (103 to 126) *	-0.23	-0.19	-0.35 (0.042)	-0.24
Barthel	20 [9 to 20] *	20 (9 to 20) *	-0.06	-0.24	-0.14	-0.06
CAMBS – disability	2 [1 to 4] *	2 (1 to 4) *	0.11	0.33 (0.021)	0.26	0.12
Ambulation Index	2 [0 to 9] *	2 (0 to 9) *	0.17	80.31	0.49 (0.034)	0.18
Handicap measures						
CAMBS – Handicap	2 [1 to 4] *	2 (1 to 4) *	0.38 (0.009)	-0.20	0.28	0.38 (0.009)
London Handicap Scale	61 [16] **	61 (16) **	-0.09	-0.08	-0.16	-0.09
Quality of Life measures						
EuroQol VAS	72 [22] **	72 (22) **	-0.39 (0.006)	0.26	-0.13	-0.39 (0.007)
SF 36:						
Physical functioning	20 [6] **	20 (6) **	-0.16	0.16	-0.32	-0.16
Physical role limitation	50 [42] **	50 (42) **	-0.18	-0.11	-0.06	-0.16
Emotional role limitation	69 [42] **	69 (42) **	-0.34 (0.021)	-0.05	-0.07	-0.34 (0.023)
Social functioning	62 [26] **	62 (26) **	-0.17	-0.03	0.11	-0.17
Mental health	72 [19] **	72 (19) **	-0.28 (0.043)	-0.09	-0.32	-0.29 (0.045)
Vitality	42 [22] **	42 (22) **	-0.22	0.03	-0.13	-0.21
Bodily pain	76 [25] **	76 (25) **	-0.07	-0.05	0.18	-0.07
General health perception	48 [25] **	48 (25) **	-0.25	0.24	-0.35	-0.25

* Median (range); ** Mean (SD); figures in parenthesis represent the statistically significant *p* values

B). Correlation with disability outcome measures

Modest correlations were found between T1-weighted total and cerebral lesion loads and the cognitive disability domain of the GNDS-R (Table 6.9); T1- and T2-weighted brain stem lesion load and the swallowing and the upper limb disability domains and the sum score of the GNDS-R (Tables 6.8 and 6.9) and the disability domain of the CAMBS (Tables 7.6 and 7.7); T1- and T2-weighted cerebellar lesion load and the upper limb disability domain of the GNDS-R (Tables 7.8 and 7.9), the FIM sum score and the AI (Tables 7.6 and 7.7); and between T2-weignted cerebellar lesion load and the bladder disability domain of the GNDS-R (Table 6.8).

C). Correlation with handicap and health related quality of life outcome measures

Modest correlations were found between T1-weighted total and cerebral lesion loads and the handicap domain of the CAMBS, the EuroQol VAS, the emotional role limitation, and mental health domain of the SF-36 (Table 7.9).

Sub-group analysis of patients with relapsing-remitting and secondary progressive MS showed similar results.

Table 7.8 Correlation between T2-lesion loads and the GNDS-R

Scale item	Score *	Correlation with MRI lesion loads			
		Cerebral	Brain Stem	Cerebellar	Total lesion load
Cognitive disability	0 [0 to 3]	0.24	0.34	-0.09	0.25
Mood disability	0 [0 to 4]	-0.04	0.17	0.48	-0.03
Visual disability	0 [0 to 3]	0.15	0.22	0.26	0.17
Speech disability	0 [0 to 3]	0.12	0.28	0.18	0.13
Swallowing disability	0 [0 to 2]	0.18	0.51 (0.003)	0.04	0.20
Upper limb disability	1 [0 to 4]	0.17	0.39 (0.003)	0.49 (0.009)	0.19
Lower limb disability	2 [0 to 4]	-0.01	0.32	0.32	0.01
Bladder disability	2 [0 to 4]	0.04	-0.01	0.23	0.04
Bowel disability	0 [0 to 5]	-0.01	0.01	0.06	-0.01
Fatigue	2 [0 to 4]	0.11	0.28	0.27	0.11
Sexual disability	0 [0 to 5]	0.16	0.21	0.17	0.16
Other disabilities	1 [0 to 4]	-0.05	0.21	0.17	-0.04

* Median (range); figures in parenthesis represent the statistically significant *p* values

Table 7.9 Correlation between T1-lesion loads and the GNDS

Scale item	Score *	Correlation with MRI lesion loads			
		Cerebral	Brain Stem	Cerebellar	Total lesion load
Cognitive disability	0 [0 to 3]	0.32 (0.028)	0.23	-0.03	0.32 (0.029)
Mood disability	0 [0 to 4]	-0.02	0.01	-0.07	-0.01
Visual disability	0 [0 to 3]	0.22	-0.28	0.23	0.24
Speech disability	0 [0 to 3]	0.21	0.01	0.25	0.21
Swallowing disability	0 [0 to 2]	0.29 (0.043)	0.39 (0.042)	0.25	0.32 (0.031)
Upper limb disability	1 [0 to 4]	0.23	0.32 (0.035)	0.34 (0.031)	0.24
Lower limb disability	2 [0 to 4]	0.08	-0.31	0.34	0.09
Bladder disability	2 [0 to 4]	0.09	0.39	0.33	0.09
Bowel disability	0 [0 to 5]	0.26	0.14	0.02	0.03
Fatigue	2 [0 to 4]	0.16	0.15	0.11	0.15
Sexual disability	0 [0 to 5]	0.22	-0.11	0.32	0.22
Other disabilities	1 [0 to 4]	-0.04	-0.17	0.23	-0.03

* Median (range); figures in parenthesis represent the statistically significant *p* values

7.9 Discussion

Mean T2- and T1-weighted lesion volumes in this study were similar to the previously reported figures from other large cohorts with similar patients' characteristics indicating that the study cohort constituted an unbiased sample from the general multiple sclerosis population (Truyen et al., 1996; Simon et al., 1998). As reported in the literature, I found the T2-weighted lesion loads to be higher than the T1-weighted lesion loads in the three anatomical sites (van Walderveen et al., 1995; Truyen et al., 1996). Similar to the report by Baumhefner and co-workers (Baumhefner et al., 1990), I also found the cerebral T2- and T2-weighted lesions to comprise the majority of the total lesion load, and the brain stem lesion load to be higher than the cerebellar lesion load on T2-weighted imaging. By contrast, the cerebellar lesion load was higher than brain stem lesion load on T1-weighted imaging. There are no published reports addressing infratentoral T1-weighted lesion loads.

7.9.1 Reliability of analysis

Although inter- and intra-rater reliability of quantitative MRI analysis has been assessed by many workers (Miller et al., 1998), there is only one published study assessing intra-rater reliability of both T2- and T1-weighted lesion load which used a methodology similar to one used in my study (van Walderveen et al., 1995). As in my study, the investigator in the van Walderveen study identified all multiple sclerosis lesions themselves on electronic data before segmenting them, and their reliability figures therefore included the reliability of both lesion recognition and delineation. In contrast, multiple sclerosis lesions in other studies of higher reliability figures were identified for the investigators on hard copies by expert neuroradiologists, and the reliability figures in such studies therefore only address the delineation process and exclude any variability related to lesion identification.

As in my study, van Walderveen and co-workers found the intra-rater variability of lesion load quantification to depend on the MRI acquisition and the extent of the abnormalities, being smaller for larger lesion loads on T2-weighted images (1.8% to 6.1%) than for smaller lesion loads on T1-weighted images (2.0% to 10.3%) (van Walderveen et al., 1995). Mean T1- and T2- weighted lesion loads and intra-rater variability in the van Walderveen study and my study were similar. However brain stem and cerebellar lesion loads in the van Walderveen study were much higher than my study, and this may explain, at least partly, the low reliability of lesion load estimation in these two regions of interest obtained in my study. Infratentorial lesions in multiple sclerosis are usually very small and difficult to recognise and delineate which is partly related to the presence of disturbing flow artefacts from the third and fourth ventricles and to the relatively poor contrast of these lesions due to the cerebellum not being in the centre of the coil (Miller et al., 1998).

Intra-rater variability of lesion quantification in the van Walderveen study fell well below their median annual increase of lesion load, which was 9% on the T2-weighted and 40% on the T1-weighted images (van Walderveen et al., 1995). With the exception of the brain stem and cerebellar T1-weighted lesion volumes, these figures are well above my intra-rater variability values suggesting that my analysis is sufficiently reliable not only for cross-sectional but also for longitudinal studies.

7.9.2 Clinical correlates

The most consistent correlations were found between T1- and T2-weighted lesion loads and the different measure of impairment particularly the cerebellar and the brain stem Functional System scores and the SNRS sum scores. These correlations were particularly high when T1- and T2-weighted brain stem lesion loads were correlated with the brain stem Functional System scores, and the T1- and T2-weighted cerebellar lesion loads correlated with the cerebellar Functional System scores. A modest correlation was also found between T1-weighted (but not T2-weighted) total and cerebral lesion loads and the mental Functional System scores. Modest correlations were also found between T1-weighted cerebellar lesion load and the EDSS which reflect the effect of cerebellar dysfunction on this ambulation biased scale.

Modest correlations were also found between the T1- and T2-weighted brain stem lesion loads and various disability measures including the swallowing, upper limb, and GNDS-R sum scores and the disability domain of the CAMBS; and between the T1- and T2-weighted cerebellar lesion loads and the upper limb disability domain of the GNDS-R, the FIM sum score, and the AI reflecting the effect of cerebellar dysfunction on these disabilities. A weak correlation was also found between T1-weighted total and cerebral lesion loads and cognitive disability domain of the GNDS-R.

Modest correlations were also found between T1-weighted total and cerebral lesion loads and overall health related quality of life as assessed by the EuroQol, and the emotional role limitation and the mental heath components of the SF-36. Interestingly the handicap domain of the CAMBS, but not the London Handicap Scale, showed similar degrees of correlation with T1-weighted total and cerebral lesion loads as the EuroQol suggesting that this scale is tapping health related quality of life rather than handicap. The significance of the observed correlations between bladder and bowel impairment and disability measures and T1- and T2- cerebellar lesion load remains unclear.

These results are compatible with the ICIDH model which predicts a high correlation between pathology and impairment but a weak correlation between pathology and disability or between pathology and handicap, and with the previously published reports of modest correlations between conventional MRI

parameters and the various impairment and disability clinical outcome measures. These results also support the previously reported high correlations between lesions at specific sites and relevant impairment and disability measures as discussed earlier. The interaction between the site of the pathological process and its severity has recently been addressed in an MRI spectroscopy study in which N-acetyl-aspartate resonance intensity in the cerebellum was found to be normal in non-ataxic multiple sclerosis patients but reduced in ataxic patients to the levels present in patients with autosomal dominant spino-cerebellar degeneration (Davie et al., 1995). Although some of the correlations between the various clinical measures and the T1-weighted lesion loads were slightly higher or more significant than with the T2-weighted lesion loads, all the observed correlations were only mild or moderate.

In discussing the complex relationship between MRI and clinical disablement and health related quality of life measures, several factors need to be considered (Miller et al., 1998):

A). Clinical outcome measures

Existing disablement and health related quality of life measures have important limitations. Correlations with more detailed and specific outcomes, such as the neuropsychological measures, have achieved higher scores.

B). Lesion size

The size of individual MS lesions may determine the nature and evolution of functional deficits. For instance, a large lesion in a fibre tract may cause an acute relapse while many small lesions developing over time may cause a progressive deficit.

C). Lesion site

Some sites in the central nervous system will when damaged by lesions have more obvious disabling consequences than others. For instance cortical lesions may have little effect on physical disability whereas lesions in eloquent sites (brain stem and the spinal cord) can have a devastating impact on the person's mobility.

D). Lesion severity

The pathological severity (demyelination or axonal loss) of lesions is crucial in determining the ultimate outcome in terms of recovery.

E). Normal appearing white matter

Although normal in appearance, white matter may have quantitative MR deficits due to microscopic pathology which contribute to disability without being abnormal on conventional MR.

Until these issues are resolved clinical and MRI measures of disease activity and burden will remain divergent. The ultimate goal of further research should not only concentrate on refining clinical and MRI correlations but also on investigating the correlations between MRI and the pathological processes within lesions. Clinical and MRI parameters should both be thought of as surrogates of the underlying disease pathology.

7.10 Conclusions

This study has confirmed the presence of a modest relation between the extent of pathology in multiple sclerosis as measured by brain lesion load on T1- and T2-weighted MRI and some clinical measures of impairment, disability and health related quality of life but not handicap. The strongest correlations were observed between brain stem and cerebellar lesion loads and the corresponding impairment measures. Although the correlation between the various clinical measures and the T1-weighted lesion loads were slightly higher than with T2-weighted lesion loads, both were only mild to moderate.

Chapter 8

CONCLUSION

Outcome measures allow the classification of patients according to the presence and severity of the disease process (pathology), the resulting disablement in terms of the clinical condition (impairment), functional capacity (disability / activities limitation), and social disadvantage (handicap / participation restriction), or according to the subjectively perceived health related quality of life. The choice of a specific outcome measure for a clinical trial depends on the nature of the study and the research hypothesis being tested. In phase II studies, which are designed to assess the biological effects of therapeutic interventions on patients, a measure of pathology or impairment would be appropriate. In phase III studies, which are designed to assess the clinical effect of therapeutic interventions on the functional capacity of patients, a measure of disability, handicap, or health related quality of life would be more desirable. Regardless of their conceptual nature, the practical value of any outcome measure depends on its clinical usefulness in terms of acceptability, ease of administration, and cost effectiveness, and on its scientific integrity in terms of reliability, validity, and responsiveness.

Multiple sclerosis is the most common cause of chronic neurological disability in young adults. It often begins with a relapsing and remitting course which progresses subsequently into a progressive phase with a gradual accumulation of wide ranging disabilities resulting in a major burden of suffering for patients and their families and substantial demands on health and social services. Many rating scales have been devised to assess the effect of experimental interventions on this illness. The scales most commonly used are the Kurtzke's Expanded Disability Status Scale, the Scripps Neurological Rating Scale, the Functional Independence Measure, the Ambulation Index, and the Cambridge Multiple Sclerosis Basic Score. The psychometric properties of these scales were investigated in this thesis and none was found to have satisfied all the

requirements of an ideal outcome measure although all had some desirable properties. Furthermore a postal survey showed that the majority of 49 leading international neurologists involved with multiple sclerosis research felt that currently existing outcome measures for multiple sclerosis are inadequate, and that there is a need for a new measure which should be patient orientated, multidimensional, and not biased towards any particular disability.

To fulfil this need, the Guy's Neurological Disability Scale was subsequently devised as a simple and user-friendly clinical disability scale capable of embracing the whole range of disabilities which could be encountered in the course of this illness in 12 separate categories which include cognition, mood, vision, speech, swallowing, upper limb function, lower limb function, bladder function, bowel function, sexual function, fatigue, and 'other disabilities'. In our hands, the Guy's Neurological Disability Scale was found to be reliable, responsive, and valid as a measure of disability. The scale was also found to be acceptable to neurologists and patients, and valid when applied by non-neurologists, over the telephone, or via a postal questionnaire. The results in this thesis lacks supportive psychometric data on a naive patient sample with more widespread disability by independent investigators. Such data will be provided by a study which now underway in four centres throughout the UK.

The correlation between pathology on one hand and impairment, disability and handicap on the other was also investigated in this thesis by assessing the correlation between T1- and T2-weighted MRI lesion load as a measure of pathology and various impairment, disability, handicap, and health related quality of life outcome measures. As predicted by the ICIDH, this study showed modest relation between the extent of pathology and clinical measures of impairment, disability and health related quality of life but not handicap. However, the strongest correlations were observed between brain stem and cerebellar lesion loads and the corresponding impairment measures.

The data discussed in this thesis have a number of limitations in relation to the reliability of the methodology and the generalisability of the results.

The reproducibility methods used in this thesis were sub-optimal in some respects. The various scales assessed were applied repeatedly on the same sample of patients by the same raters and the possibility of patients' and raters' bias cannot be excluded. However, as discussed in chapter 5 and chapter 6, all the

raters were blinded to their own and other raters' previous scores, and open discussions about patients' clinical conditions were avoided amongst themselves. In the inter-rater reliability studies, patients were assessed independently by the two raters and no fixed order for the examination was observed so as to reduce the effect of patients' bias which might have resulted from practice effect or fatigue. Data for the intra-rater reliability and responsiveness studies were collected at three monthly intervals so as to reduce raters' and patients' bias which might have resulted from recall of the previous assessments. The effect of this potential source of bias is unlikely to have been significant since the inter-rater reliability figures were often higher than the intra-rater reliability figures. It is also unlikely that the familiarity of the patients to the assessors or the frequent administrations of the Guy's Neurological Disability Scale have biased the results significantly since the reliability figures of the various methods of scale administration were often lower then the initial inter-rater reliability figures which were obtained when the scale was administered by the two raters for the first time. It is also worth noting that the repeated administration of this scale did not result in any significant 'regression to the mean' of the various sub-and sum scores as the median (range) scores remained relatively stable throughout the repeated scale administration during the second phase of the study.

The number patients recruited for this study was small. Sample size estimation was not done *a priori* since I was constrained by the number of patients in the ongoing interferon beta 1a trials. Although this stricture might have resulted in a degree of type II statistical error (false negative results), the confidence intervals of the various reliability results were relatively narrow indicating that the true reliability coefficients were reasonably close to those estimates in this study. Streiner and Norman (1995) proposed that sample sizes should vary according to the magnitude of the expected reliability coefficients and the width of the confidence intervals. They suggested that for confidence intervals of $\pm$ 0.10 and an α of 0.05 a sample of 50 and 35 patients will result in a "statistical overkill" for any reliability coefficients over 0.85 and 0.90 respectively.

The measurement properties of the various scales have also been examined in a selected sample of highly motivated patients with mild to moderate degree of

disability (EDSS 0-7.5). Ideally such a study should have been conducted in a population based sample from a geographically defined area to ensure the accurate representation of the patients in the general MS population. Natural history studies suggest that only 78-85% of patients have EDSS scores <8.0 (Goodkin et al., 1989; Swingler and Compston, 1992; Rodriguez et al., 1994; Midgard et al., 1996). As the sample in this study was restricted, the results discussed in this thesis should only be generalised to patients with EDSS <8.0.

The majority of the Guy's Neurological Disability sub-scales in the studied cohort were skewed to the less disabled end of the scale suggesting a 'floor' effect. This skewness is likely to have been related to the nature of disability in this sample rather than the scale itself since the frequency distribution of these sub-scales was near Normal in the postal survey data set. Although such a strong 'floor' effect limits the legitimate use of parametric statistics, particularly factor and regression analysis, it does not affect the results of non-parametric statistics or kappa coefficients. The use of parametric statistics for the calculation of mean score differences and repeatability coefficients remains appropriate since score differences are likely to have Normal distribution (Bland and Altman, 1986). Furthermore, the Guy's Neurological Disability Scale was developed as a multi-item rather than a single item measure and the frequency distribution of its sum scores was Normal.

Multi-item rating scales are known have superior measurement properties to single item measures. This is not surprising since single item measures are conceptually unlikely to fully represent complex theoretical concepts such as disability (Nunnally, 1978). Furthermore single items are unable to make the fine differentiations between patients which are desirable for most measurement problems. Perhaps the most important limitation of single item measure is the fact that their measurement properties, particularly validity, are difficult to examine (Mclever and Carmines, 1981).

The assessment of responsiveness was also sub-optimal in this study. As discussed in chapter 3, the notion of responsiveness was originally based on the simple construct that the goal of therapy is to induce change in health status (Norman et al., 1997). A number of measures of responsiveness based on the measurement of change following therapeutic interventions, including Student-t test, Wilcoxon Signed Rank test and effect size, have been proposed. Although

the computation of these methods is straightforward, this approach has some theoretical limitations since the magnitude of the resulting coefficients is dependent on the actual change induced by treatment and cannot therefore be viewed as the sole property of the instrument itself. As a consequence some investigators have abandoned this strategy and opted to estimate responsiveness independent of any particular intervention by identifying those patients who have and have not changed over time to a clinically important extent using other criteria of change, such as patient or physician perception of change (Guyatt et al., 1987). This strategy was adopted in this thesis. Unfortunately the reliability and validity of this method have never been established (Norman et al., 1997). The judgement of change is psychologically difficult. Patients must be able to quantify their present and initial status and then perform a mental subtraction. There is evidence that patients are in fact unable to recall their initial health status and their judgements are usually biased towards their present status (Ross, 1989). The use of clinician judgement of change is unlikely to have avoided this bias since I had to use the patients as informants and the same confounding might have occurred.

The patients in this study were also taking part in a therapeutic trial of interferon beta 1a, and such treatment might have affected their clinical condition and shaped their appraisal of their clinical status. However this effect is unlikely to have been important since all patients were blinded as to the treatment and the study was done under common overall treatment effect.

Finally, although the Guy's Neurological Disability Scale was devised as a disability scale, its sum scores and many of its sub-scales correlated highly not only with other disability measures but also with many impairment measures suggesting that the scale is tapping a combination of impairment and disability. This is likely to have been due to the nature of the scale itself and to the conceptual difficulties in separating impairment from disability within the ICIDH model. The three basic constructs in this model describe closely related theoretical entities which overlap considerably. The distinction between impairment and disability is particularly difficult as evident in the mood, visual, speech, swallowing, bladder, and bowel disability Guy's Neurological Disability sub-scales. These 'disabilities' are graded according to the presence and severity of the relevant symptoms, i.e. according to the subjective degree of 'impairment'. Visual 'impairment' on the other hand is traditionally assessed by measuring

patients' 'disability' in reading various visual charts (Hughes and Sharrack, 1996). This imperfection of the ICIDH model explains the surprisingly high correlation between the Scripps Neurological Rating Scale as a measure of impairment and the Functional Independence Measure, the disability domain of the Cambridge Multiple Sclerosis Basic Score, and the physical functioning domain of the Short Form 36 which has been discussed in chapter 5. Despite these difficulties, the ICIDH model remains the best framework currently available for health science research.

Provided that the currently ongoing additional study on the Guy's Neurological Disability Scale shows that the results discussed in this thesis can be generalised, this scale will be a helpful new clinical outcome measure for multiple sclerosis which is capable of providing relevant information on the wide range of disability as experienced by patients. Additional assessment of impairment, handicap, and health related quality of life would be necessary to provide a more comprehensive appraisal which incorporates the objective and the subjective aspects of patients' health status.

Appendix 1

THE REVISED GUY'S NEUROLOGICAL DISABILITY SCALE

1). Cognitive Disability

A. Interview:

Do you have any problems with your memory or your ability to concentrate and work things out?

☐ yes ☐ no

Do your family or friends think that you have such a problem?

☐ yes ☐ no

If the answer to either question is 'yes':

Do you need help from other people for planing your normal daily affairs, handling money or making decisions?

☐ yes ☐ no

If 'yes': (To the examiner)

Is the patient orientated in time, place and person?

☐ yes, fully
☐ yes, partially *
☐ no, totally disorientated *

** If the patient is not fully orientated, all their answers should be verified by the main carer(s) whose answers should take precedence.*

B)-Scoring:

0- No cognitive problems.
1- Cognitive problems not noticeable to family or friends.
2- Cognitive problems noticeable to family or friends but not requiring help from others.
3- Cognitive problems requiring help from others for normal daily affairs; patient is fully orientated in time, place and person.
4- Cognitive problems requiring help from others for normal daily affairs; patient is not fully orientated.
5- Patient is completely disorientated in time, place and person.

2). Mood Disability

A)-Interview:

Have you been feeling anxious, irritable, depressed, or had any mood swings during the last month?

☐ yes ☐ no

Are you taking any medications for such problems?

☐ yes ☐ no

If the answer to the first question is 'yes':

Has this problem affected your ability to do any of your usual daily activities such as work, housework, or normal social activities with family and friends?

☐ yes ☐ no

If 'yes':

Has this problem been severe enough to prevent you from doing all your usual activities?

☐ yes ☐ no

Have you been admitted to hospital for treatment of your mood problem during the last month?

☐ yes ☐ no

B. Scoring:

0- No mood problems.
1- Asymptomatic on current drug treatment.
2- Mood problems present but not affecting the patient's ability to perform any of their usual daily activities.
3- Mood problems affecting the patient's ability to perform some of their usual daily activities.
4- Mood problems preventing the patients from doing all their usual daily activities.
5- Mood problems requiring inpatient management.
X- Unknown, please score as the mean of the cognitive and fatigue disability scores rounded to the nearest integer (see results of factor analysis).

3). Visual disability

A. Interview:

Do you have any problems with your vision which can't be corrected with ordinary glasses?

☐ yes ☐ no

If 'yes':

Can you read ordinary newspaper print (with ordinary glasses if worn, but not magnifying lenses)?

☐ yes ☐ no

If 'no':

Can you read large newspaper print?

☐ yes ☐ no

If 'no':

Can you count your fingers if you hold your hand out in front of you?

☐ yes ☐ no

If 'no':

Can you see your hand if you move it in front of you?

☐ yes ☐ no

B. Scoring:

0- No visual problems.
1- Visual problems (blurred vision, diplopia, scotomas) but patient is still able to read ordinary newspaper print.
2- Unable to read ordinary newspaper print.
3- Unable to read large newspaper print.
4- Unable to count fingers if they hold their hand out in front of them.
5- Unable to see hand movement if they move their hand in front of them.

4). Speech disability

A. Interview

Do you have any problems with your speech?

☐ yes ☐ no

If 'yes':

Do you have to repeat yourself when speaking to strangers?

☐ yes ☐ no

If 'yes':

Do you have to repeat yourself when speaking to your family or close friends?

☐ yes ☐ no

If 'yes':

Do you need to use sign language, or the help of your carer to make people understand you?

☐ yes ☐ no

If 'yes': (to the examiner)

Is the patient able to communicate effectively using these methods?

☐ yes ☐ no

B. Scoring:

0- No speech problems.
1- Speech problems which does not require the patient to repeat themselves when speaking to strangers.
2- Speech problems which require the patient to repeat themselves when speaking to strangers.
3- Speech problems which require the patient to repeat themselves when speaking to their family and close friends.
4- Speech problems making speech difficult to understand; patient is able to communicate effectively by using sign language or the help of their carers.
5-Speech problems making speech difficult to understand, patient is unable to communicate effectively by using sign language or the help of their carers.

5). Swallowing disability

A. Interview:

Do you have to take care when swallowing solids or fluids?

☐ yes ☐ no

If 'yes':

Do you have to take care when swallowing with most meals?

☐ yes ☐ no

If 'yes':

Do you need a special diet such as soft or liquidated food to help with your swallowing?

☐ yes ☐ no

If 'yes':

Do you choke with most meals?

☐ yes ☐ no

If yes:

Do you have a feeding tube (nasogastric or gastrostomy tube)?

☐ yes ☐ no

B)-Scoring:

0- No swallowing problems.
1- Needs to be careful when swallowing solids or liquids but not with most meals.
2- Needs to be careful when swallowing solids or liquids with most meals; patient is able to eat food of normal consistency.
3- Needs specially prepared food of modified consistency.
4- Tendency to choke with most meal.
5- Dysphagia requiring nasogastric or gastrostomy tube.

6). Upper limb disability

A. Interview:

Do you have any problems with your hands or arms?

☐ yes ☐ no

If 'yes':

Do you have any difficulty in doing any of your zips or buttons?

☐ yes ☐ no

If 'yes':

Are you able to do all of your zips and buttons without help?

☐ yes ☐ no

Do you have any difficulty in tying a bow in laces or strings?

☐ yes ☐ no

If 'yes':

Are you able to tie a bow in laces or strings without help?

☐ yes ☐ no

Do you have any difficulty washing and brushing your hair?

☐ yes ☐ no

If 'yes':

Are you able to wash and brush your hair without help?

☐ yes ☐ no

Do you have any difficulty feeding yourself?

☐ yes ☐ no

If 'yes':

Are you able to feed yourself without help?

☐ yes ☐ no

If unable to do any of the functions listed:

Can you use your hands or arms for any other function?

☐ yes ☐ no

B. Scoring:

0- No upper limb problems.
1- Problems in one or both arms, not affecting the ability to do any of the functions listed.
2- Problems in one or both arms, affecting some but not preventing any of the functions listed.
3- Problems in one or both arms, affecting all or preventing one or two of the functions listed.
4- Problems in one or both arms preventing three or all of the functions listed.
5- Unable to use either arm for any purposeful movements.

7). Lower limb disability

A. Interview:

Do you have any problems with your walking?

☐ yes ☐ no

If 'yes':

Do you use a walking aid?

☐ yes ☐ no

If 'yes':

A. How do you usually get around outdoors?

☐ without aid

OR ☐ with one stick or crutch or holding to someone's arm

OR ☐ with two sticks or crutches or one stick or crutch and holding to someone's arm

OR ☐ with a wheelchair

B. How do you usually get around indoors?

☐ without aid

OR ☐ with one stick or crutch or holding to someone's arm

OR ☐ with two sticks or crutches or one stick or crutch and holding to someone's arm

OR ☐ with a wheelchair

If you use a wheelchair:

Can you stand and walk few steps with help?

☐ yes ☐ no

B. Scoring:

0- Walking is not affected.
1- Walking is affected but patient is able to walk independently.
2- Usually uses unilateral support (single stick or crutch, one arm) to walk outdoors, but walks independently indoors.
3- Usually uses bilateral support (two sticks or crutches, frame, or two arms) to walk outdoors, **or** unilateral support (single stick or crutch, or one arm) to walk indoors.
4- Usually uses wheelchair to travel outdoors, **or** bilateral support (two sticks or crutches, frame, or two arms) to walk indoors.
5- Usually uses a wheelchair indoors.

8). Bladder disability

A. Interview

Do you have any problems with your bladder?

☐ yes ☐ no

Are you taking any medications for such problems?

☐ yes ☐ no

If the answer to the first question is 'yes':

Do you have to rush to the toilet, go frequently, or have difficulty in starting to pass urine?

☐ yes ☐ no

Have you been incontinent last month?

☐ yes ☐ no

If 'yes':

Have you been incontinent last week?

☐ yes ☐ no

If 'yes':

Have you been incontinent every day?

☐ yes ☐ no

Do you need to use a catheter to empty your bladder?

☐ yes ☐ no

Do you need a permanent catheter in the bladder, or (for men only) do you use a sheath to collect your urine?

☐ yes ☐ no

B. Scoring:

0- Normal bladder problems.
1- Aasymptomatic on current drug treatment.
2- Urinary frequency, urgency, or hesitancy with no incontinence.
3- Occasional urinary incontinence (once or more during the last month but not every week) **or** intermittent catheterisation without incontinence.
4- Frequent urinary incontinence (once a week or more during the last month but not daily), **or** occasional urinary incontinence despite regular intermittent catheterisation.
5- Daily urinary incontinence or permanent catheter (urethral / suprapubic) or penile sheath.

9). Bowel disability

A. Interview:

Do you have any problems with your bowel movements?

☐ yes ☐ no

Are you on any medicines for such problems?

☐ yes ☐ no

If the answer to the first question is 'yes':

Do you suffer with constipation?

☐ yes ☐ no

If 'yes':

Do you need to take any laxatives or use suppositories for this?

☐ yes ☐ no

Do you usually use enemas?

☐ yes ☐ no

Do you usually evacuate your stools manually?

☐ yes ☐ no

Do you have to rush to the toilet to open your bowels?

☐ yes ☐ no

Have you had any bowel accidents (been incontinent of faeces) last week?

☐ yes ☐ no

If 'yes':

Have you had bowel accidents every week?

☐ yes ☐ no

B)-Scoring:

0- No bowel problems.

1- Asymptomatic on current drug treatment or constipation not requiring any treatment.

2- Constipation requiring laxatives or suppositories **or** faecal urgency.

3- Constipation requiring the use of enemas.

4- Constipation requiring manual evacuation of stools **or** occasional faecal incontinence (once or more during the last month but not every week).

5- Weekly faecal incontinence.

10). Fatigue disability

A. Interview:

Have you been feeling tired or getting tired easily during the last month?

☐ yes ☐ no

If 'yes':

Have you been feeling tired most days?

☐ yes ☐ no

Has this tiredness affected your ability to do any of your usual activities such as work, housework, or normal social activities with family and friends?

☐ yes ☐ no

If 'yes':

Has this tiredness been severe enough o prevent you from doing all of your usual activities?

☐ yes ☐ no

If 'yes':

Has the tiredness been severe enough to prevent you from doing all physical activities?

☐ yes ☐ no

B. Scoring:

0- Absent.
1- Occasional fatigue present some days.
2- Frequent fatigue present most days.
3- Fatigue affecting the patient's ability to perform some of their usual daily activities.
4- Fatigue preventing the patient from doing all their usual daily activities.
5- Fatigue preventing the patient from doing all physical activities.
X- Unknown, please score as the mean of the cognitive and mood disability scores rounded to the nearest integer (see results of factor analysis).

11). Sexual disability

A. Interview:

The next set of questions relates sexual function. Do you mind if I ask you about this?

☐ yes
☐ no
☐ not applicable (celibate)

If the patient agrees:

Do you have any problems in relation to your sexual function?

☐ yes ☐ no

If 'yes':

Do you suffer with lack of sexual interest?

☐ yes ☐ no

Do you have any problems satisfying yourself or your sexual partner?

☐ yes ☐ no

Is your sexual function affected by any physical problem such as altered genital sensation, pain, or spasms?

☐ yes ☐ no

Do you have any problems with:

(for men): erection / ejaculation?
(for women): vaginal lubrication / orgasm?

☐ yes ☐ no

If physical or sexual problems are present:

Do any of these difficulties totally prevent your sexual activities?

☐ yes ☐ no

B. Scoring:

0- Normal sexual function **or** persons who are voluntarily celibate.
1- Reduced sexual interest.
2- Problems satisfying oneself or sexual partner.
3- Physical problems interfering but not preventing sexual function.
4- Autonomic problems interfering but not preventing sexual function.
5- Physical or autonomic problems totally preventing sexual function.
X- Unknown, please score as the mean of the lower limb, bladder, and bowel disability scores rounded to the nearest integer (see results of factor analysis).

12). Other disabilities

A. Interview:

Do you have other problems due to MS such as pain, spasms, or dizziness which have not been mentioned so far?

☐ yes ☐ no

Are you taking any medicines for such problems?

☐ yes ☐ no

If the answer to either question is 'yes':

Please name your worst problem: .....

Has this problem affected your ability to do any of your usual daily activities?

☐ yes ☐ no

Has this problem been severe enough to prevent you from doing all your usual daily activities?

☐ yes ☐ no

Have you been admitted to hospital for treatment of this problem?

☐ yes ☐ no

B. Scoring:

0- Absent.
1- Asymptomatic on current drug treatment.
2- Problems present, but are not affecting the patient's ability to perform any of their usual daily activities.
3- Problems affecting the patient's ability to perform some of their usual daily activities.
4- Problems preventing the patient from doing all their usual daily activities.
5- Problems requiring hospital admission for assessment or treatment.

Appendix 2

THE POSTAL GUY'S NEUROLOGICAL DISABILITY SCALE

A. About your MS

1) Memory and Concentration

			YES	NO
1.1	a.	Do you have any problems with your memory or your ability to concentrate and work things out?	☐	☐
	b.	Do your family or friends think that you have such a problem?	☐	☐

If you answered "No" to *both* previous questions, go straight to question 2.1.

If you answered "Yes" to *either* of the previous questions:

1.2	a.	Do you need help from other people for planing your daily affairs, handling money or making decisions?	☐	☐
	b.	**If you answered "Yes" to 1.2a,** have you been feeling confused and not sure where you are or what time of the day, week, or month it is?	☐	☐

2) Mood and Anxiety

			YES	NO
2.1	a.	Have you been feeling anxious, irritable, depressed, or had any mood swings during the last month?	☐	☐
	b.	Are you taking any medication for such problem	☐	☐

If you answered "No" to *both* previous questions, go straight to question 3.1.

If you answered "Yes" to 2.1a:

2.2	a.	Has this problem affected your ability to do any of your usual daily activities such as work, housework, or normal social activities with family and friends?	☐	☐
	b.	**If you answered "Yes" to 2.2a,** has this problem been severe enough to prevent you from doing **all** your usual activities?	☐	☐
	c.	Have you been admitted to hospital for treatment of your mood problem during the last month?	☐	☐

3) Eyes and Vision

			YES	NO
3.1		Do you have any problems with your vision that can't be corrected with ordinary glasses?	☐	☐

If you answered "No", go straight to question 4.1.

If you answered "Yes":

3.2	a.	Can you read **ordinary** newspaper print (with ordinary glasses if worn, but not magnifying lenses)?	☐	☐
	b.	**If you answered "No" to 3.2a,** can you read **large** newspaper print?	☐	☐
	c.	**If you answered "No" to 3.2b,** can you count your fingers if you hold your hand out in front of you?	☐	☐
	d.	**If you answered "No" to 3.2c,** can you see your hand if you move it in front of you?	☐	☐

4) Speaking and Communicating

			YES	NO
4.1		Do you have any problems with your speech?	☐	☐
		If you answered "No", go straight to question 5.1.		
		If you answered "Yes":		
4.2	a.	Do you have to repeat yourself when speaking to your family or close friends?	☐	☐
	b.	**If you answered "Yes" to 4.2a,** do you need to use sign language or the help of your carer to make people understand you?	☐	☐
	c.	**If you answered "Yes" to 4.2b,** are you able to communicate effectively using these methods?	☐	☐

5) Swallowing

			YES	NO
5.1.		Do you have to take care when swallowing solids or fluids?	☐	☐
		If you answered "No", go straight to question 6.1.		
		If you answered "Yes":		
5.2.	a.	Do you have to take care when swallowing with most meals?	☐	☐
	b.	**If you answered "Yes" to 5.2a,** do you need a special diet such as soft or liquidated food to help with your swallowing?	☐	☐
	c.	**If you answered "Yes" to 5.2b,** do you choke with most meals?	☐	☐
	d.	**If you answered "Yes" to 5.2c,** do you have a feeding tube (nasogastric or gastrostomy tube)?	☐	☐

6) Arms and Hands

			YES	NO
6.1		Do you have any problems with your hands or arms?	☐	☐
		If you answered "No", go straight to question 7.1.		
		If you answered "Yes",		
6.2	a.	Do you have any difficulty in doing **any** of your zips or buttons?	☐	☐
	b.	**If you answered "Yes" to 6.2a,** are you able to do **all** your zips **and** buttons without help?	☐	☐
6.3	a.	Do you have any difficulty in washing and brushing your hair?	☐	☐
	b.	**If you answered "Yes" to 6.3a,** are you able to wash **and** brush your hair without help?	☐	☐
6.4	a.	Do you have any difficulties in tying a bow in laces or strings?	☐	☐
	b.	**If you answered "Yes" to 6.4a,** are you able to tie a bow in laces or strings without help?	☐	☐
6.5	a.	Do you have any difficulty in feeding yourself	☐	☐
	b.	**If you answered "Yes" to 6.5a,** are you able to feed yourself without help?	☐	☐
6.6		**If unable to do *any* of the functions listed in Question 6 above,** are you able to use your hands or arms for any other function?	☐	☐

7) Legs

		YES	NO
7.1	Do you have any problems with your walking?	☐	☐
	If you answered "No", go straight to question 8.1.		
	If you answered "Yes":		

7.2 a. Do you use a walking aid? ☐ ☐

b. **If you answered "Yes" to 7.2a,** how do you **usually** get around outdoors? (please tick only one of the 4 boxes below)

- Without aid ☐
- With one stick or crutch **or** holding someone's arm ☐
- With two sticks or crutches, a frame, **or** one stick or crutch **and** holding someone's arm ☐
- With a wheelchair ☐

c. How do you **usually** get around indoors? (please tick only one of the 4 boxes below)

- Without aid ☐
- With one stick or crutch **or** holding someone's arm ☐
- With two sticks or crutches, a frame, **or** one stick or crutch **and** holding someone's arm ☐
- With a wheelchair ☐

d. **If you answered "With a wheelchair",** can you stand and walk a few steps with help? ☐ ☐

8) <u>Going to the toilet (bladder)</u>

		YES	NO
8.1.	a. Do you have any problems with your bladder?	☐	☐
	b. Are you taking any medications for such problems?	☐	☐

If you answered "No" to both previous questions, go straight to question 9.1.

If you answered "Yes" to 8.1a :

8.2.	a.	Do you have to rush to the toilet?	☐	☐
	b.	Do you have to go to the toilet frequently?	☐	☐
	c.	Do you have difficulty in starting to pass urine?	☐	☐
8.3	a.	Have you been incontinent last month?	☐	☐
	b.	**If you answered "Yes" to 8.3a,** have you been incontinent in the last week?	☐	☐
	c.	**If you answered "Yes" to 8.3b,** have you been incontinent every day?	☐	☐
8.4.		Do you need to use a catheter to empty your bladder?	☐	☐
8.5.		Do you need a permanent catheter in the bladder, or **(for men only)** do you use a sheath to catch the urine?	☐	☐

9) <u>Going to the toilet (bowel)</u>

			YES	**NO**
9.1.	a.	Do you have any problems with your bowel movements?	☐	☐
	b	Are you on any treatment for such problems?	☐	☐

If you answered "No", to *both* questions, go straight to question 10.1.

If you answered "Yes" to 9.1a:

9.2.	a.	Do you suffer with constipation?	☐	☐

b. **If you answered "Yes" to 9.2a,** do you need to take laxatives or use suppositories for this? ☐ ☐

c. Do you need to use enemas for this? ☐ ☐

d. Do you have to evacuate your stools manually? ☐ ☐

9.3 Do you have to rush to the toilet to open your bowels? ☐ ☐

9.4 a. Have you had any bowel accidents (been incontinent of faeces) in the *last week?* ☐ ☐

b. **If you answered "Yes" to 9.4a,** have you had bowel accidents *every week?* ☐ ☐

10) <u>Fatigue</u>

		YES	NO
10.1	Have you been feeling tired or getting tired easily during the last month?	☐	☐

If you answered "No", go straight to question 11.1.

If you answered "Yes"

10.2	Have you been feeling tired or getting tired easily most days?	☐	☐

10.3 a. Has this tiredness affected your abilities to do **any** of your usual activities such as work or normal social activities with family and friends? ☐ ☐

b. **If you answered "Yes" to 10.3a,** has this tiredness been severe enough to prevent you from doing **all** your usual activities? ☐ ☐

c. **If you answered "Yes" to 10.3b,** has this tiredness been severe enough to confine you to bed or chair and prevent you from doing **all** physical activities? ☐ ☐

11) Sexual activity

			YES	NO
11.1		Do you have any problems in relation to your sexual function?	☐	☐
		If you answered "No", go straight to question 12.1		
		If you answered "Yes"		
11.2	a.	Do you suffer with lack of sexual interest?	☐	☐
	b.	Do you have any problems satisfying yourself or your sexual partner?	☐	☐
11.3		Is your sexual function affected by any physical problem such as altered genital sensation, pain, or spasms?	☐	☐
		Please state which:		
11.4		Do you have any problems:		
	a.	(For men): erection?	☐	☐
	b.	(For men): ejaculation?	☐	☐
	c.	(For women): vaginal lubrication?	☐	☐
	d.	(For women): orgasm?	☐	☐
11.5		**If you answered "Yes" to 11.3 - 11.4,** do any of these difficulties totally prevent your sexual activities?	☐	☐

12) Other problems

		YES	NO
12.1	Do you have any other problems due to MS such as pain, spasms, or dizziness which have not been mentioned so far?	☐	☐
12.2	**If you answered "No", go straight to part B of the questionnaire.**		
	If you answered "Yes", please name your worst problem:		
12.3 a.	Has this problem affected your ability to do **any** of your usual daily activities?	☐	☐
b.	**If you answered "Yes" to 12.3a,** has this problem been severe enough to prevent you from doing **all** your usual daily activities?	☐	☐
c.	**If you answered "Yes" to 12.3b,** have you been admitted to hospital for treatment of this problem during the last month?	☐	☐

B. About you:

1. Are you:

☐ Male

☐ Female

2. How old are you? ..... years

3. Are you currently working?

☐ Yes, full time

☐ Yes, part time

☐ No

D. Completing this questionnaire

1. Did you need help from anyone to complete this questionnaire?

☐ Yes

☐ No

2. If yes, what type of help did you need?

☐ I needed help reading the questions

☐ I needed help in writing the answers

☐ I needed help to work out how to answer the questions.

Thank you for completing this questionnaire.

Please return it in the envelope provided

BIBLIOGRAPHY

Acheson, J.F. and Sanders, M.D. (1995) Vision. *Journal of Neurology, Neurosurgery & Psychiatry* **59**, 4-15.

Amato, M.P., Fratiglioni, L., Groppi, C., Siracusa, G. and Amaducci, L. (1988) Interrater reliability in assessing functional systems and disability on the Kurtzke scale in multiple sclerosis. *Archives of Neurology* **45**, 746-748.

Applegate, W.B., Blass, J.P. and Williams, T.F. (1990) Instruments for the functional assessments of older patients. *New England Journal of Medicine* **322**, 1207-1214.

Arnold, D.L., Wolinsky, J.S., Matthews, P.M. and Falini, A. (1997) The use of magnetic resonance spectroscopy in the evaluation of the natural history of multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **64 Suppl 1**, S94-101.

Badley, E.M. (1993) An introduction to the concepts and classifications of the international classification of impairments, disabilities, and handicaps. *Disability & Rehabilitation* **15**, 161-178.

Baumhefner, R.W., Tourtellotte, W.W., Syndulko, K., Waluch, V., Ellison, G.W., Meyers, L.W., Cohen, S.N., Osborne, M. and Shapshak, P. (1990) Quantitative multiple sclerosis plaque assessment with magnetic resonance imaging. Its correlation with clinical parameters, evoked potentials, and intra-blood-brain barrier IgG synthesis. *Archives of Neurology* **47** , 19-26.

Beck, A.T., Weissman, A., Lester, D. and Trexler, L. (1974) The measurement of pessimism: the hopelessness scale. *Journal of Consulting & Clinical Psychology* **42**, 861-865.

Benton, A.L. and Hamsher, K.de.S. (1976) *Multilingual Aphasia Examination*, Iowa City, Iowa: University of Iowa.

Bergner, M., Bobbitt, R.A., Carter, W.B. and Gilson, B.S. (1981) The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* **19**, 787-805.

Bernstein, M.L., Despars, J.A., Singh, N.P., Avalos, K., Stansbury, DW and Light, R.W. (1994) Reanalysis of the 12-minute walk in patients with chronic obstructive pulmonary disease. *Chest* **105**, 163-167.

Bland, J.M. and Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307-310.

Blumhardt, L.D. and Wood, C. (1996) The economics of multiple sclerosis: a cost of illness study. *British Journal of Medical Economics* **10**, 99-118.

British and Dutch Multiple Sclerosis Azathioprine Trial Group (1988) Double-masked trial of azathioprine in multiple sclerosis. *Lancet* **2**, 179-183.

British Standards Institution (1979) *Precision of test methods I: guide for the determination and reproducibility of a standard test methods*, London: British Standards Institution.

Brosseau, L. and Wolfson, C. (1994) The inter-rater reliability and construct validity of the Functional Independence Measure for multiple sclerosis subject. *Clinical Rehabilitation* **8**, 107-115.

Bury, M.R. (1987) The ICIDH: a review of research and prospects. *International Disability Studies* **9**, 118-122.

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D. and Wallace, E.P. (1993) Development of a fatigue scale. *Journal of Psychosomatic Research* **37**, 147-153.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37-46.

Cohen, J. (1968) Weighted kappa: nominal scale agreement with precision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213-220.

Cohen, J. (1977) *Statistical power analysis for the behavioural sciences*, London: Academic Press.

Colchester, A.C., Zhao, J., Holton-Tainter, K.S., Henri, C.J., Maitland, N., Roberts, P.T., Harris, C.G. and Evans, R.J. (1996) Development and preliminary evaluation of VISLAN, a surgical planning and guidance system using intra-operative video imaging. *Medical Image Analysis* **1**, 73-90.

Compston, A. (1998) Genetic susceptibility to multiple sclerosis. In: Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews, B. and Wekerle, H., (Eds.) *McAlpine's multiple sclerosis*, 3 edn. pp. 101-144. London: Churchill Livingstone

Coughlan, A.K. and Hollows, S.E. (1985) *The Adult Memory and Information Processing Battery*, Leeds: St. James's University Hospital.

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297-334.

Davie, C.A., Barker, G.J., Webb, S., Tofts, P.S., Thompson, A.J., Harding, A.E., McDonald, W.I. and Miller, D.H. (1995) Persistent functional deficit in multiple sclerosis and autosomal dominant cerebellar ataxia is associated with axon loss. *Brain* **118**, 1583-1592.

De Stefano, N., Matthews, P.M., Antel, J.P., Preul, M., Francis, G. and Arnold, D.L. (1995) Chemical pathology of acute demyelinating lesions and its correlation with disability. *Annals of Neurology* **38**, 901-909.

Devinsky, O. (1995) Outcome research in neurology: incorporating health-related quality of life. *Annals of Neurology* **37**, 141-142.

Deyo, R.A., Diehr, P. and Patrick, D.L. (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials* **12**, 142S-158S.

Ebers, G.C. (1994) Treatment of multiple sclerosis. *Lancet* **343**, 275-279.

Ebers, G.C. and Sadovnick, A.D. (1993) The geographic distribution of multiple sclerosis: a review. *Neuroepidemiology* **12**, 1-5.

Ebers, G.C., Sadovnick, A.D. and Risch, N.J. (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* **377**, 150-151.

Ebrahim, S. (1995) Clinical and public health perspectives and applications of health-related quality of life measurement. *Social Science & Medicine* **41**, 1383-1394.

Ellison, G.W., Myers, L.W. and Leake, B.D. (1993) Responsiveness of the Disability Status Scale (DSS). *Neurology* **43** , A204 (Abstract)

Ellison, G.W., Myers, L.W., Leake, B.D., Mickey, M.R., Ke, D., Syndulko, K. and Tourtellotte, W.W. (1994) Design strategies in multiple sclerosis clinical trials. The Cyclosporine Multiple Sclerosis Study Group. *Annals of Neurology* **36 Suppl**, S108-S112

Enderby, P. (1988) The assessment of dysarthria: a challenge to more than the ears. *Clinical Rehabilitation* **2**, 267-273.

EuroQol Group (1990) EuroQol: a new facility for measurement of health related quality of life. *Health Policy* **16**, 199-208.

Ffrench-Constant, C. (1994) Pathogenesis of multiple sclerosis. *Lancet* **343**, 271-275.

Filippi, M., Horsfield, M.A., Hajnal, J.V., Narayana, P.A., Udupa, JK, Yousry, T.A. and Zijdenbos, A. (1998) Quantitative assessment of magnetic resonance imaging lesion load in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **64 Suppl 1**, S88-S93

Filippi, M., Paty, D.W., Kappos, L., Barkhof, F., Compston, D.A., Thompson, A.J., Zhao, G.J., Wiles, C.M., McDonald, W.I. and Miller, D.H. (1995) Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study. *Neurology* **45**, 255-260.

Fine, B.J. and Kobrick, J.L. (1983) Individual differences in distance estimation: comparison of judgements in the field with those from projected slides of the same scenes. *Perceptual & Motor Skills* **57**, 3-14.

Fleiss, J.L. and Schrout, P.E. (1978) Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* 259-262.

Francis, D.A., Bain, P., Swan, A.V. and Hughes, R.A. (1991) An assessment of disability rating scales used in multiple sclerosis. *Archives of Neurology* **48**, 299-301.

Franklin, G.M., Heaton, R.K., Nelson, L.M., Filley, C.M. and Seibert, C. (1988) Correlation of neuropsychological and MRI findings in chronic/progressive multiple sclerosis. *Neurology* **38**, 1826-1829.

Freeman, J.A., Langdon, D.W., Hobart, J.C. and Thompson, A.J. (1996) The impact of inpatient rehabilitation on progressive multiple sclerosis. *Annals of Neurology* **42**, 236-244.

Garratt, A.M., Ruta, D.A., Abdalla, M.I., Buckingham, J.K. and Russell, I.T. (1993) The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *British Medical Journal* **306**, 1440-1444.

Gass, A., Barker, G.J., Kidd, D., Thorpe, J.W., MacManus, D., Brennan, A., Tofts, P.S., Thompson, A.J., McDonald, W.I. and Miller, D.H. (1994) Correlation of magnetization transfer ratio with clinical disability in multiple sclerosis. *Annals of Neurology* **36**, 62-67.

Gelber, R.D., Lenderking, W.R., Cotton, D.J., Cole, B.F., Fischl, M.A., Goldhirsch, A and Testa, M.A. (1992) Quality-of-life evaluation in a clinical trial of zidovudine therapy in patients with mildly symptomatic HIV infection. *Annals of Internal Medicine* **116**, 961-966.

Gilbert, J.J. and Sadler, M. (1983) Unsuspected multiple sclerosis. *Archives of Neurology* **40**, 533-536.

Gill, T.M. (1995) Quality of life assessment: values and pitfalls. *Journal of the Royal Society of Medicine* **88**, 680-682.

Goldberg, D.P. and Williams, P. (1988) *A users guide to the General Health Questionnaire*, Windsor: NFER Nelson.

Goodkin, D.E. (1994) Interferon beta-1b. *Lancet* **344**, 1057-1060.

Goodkin, D.E., Cookfair, D., Wende, K., Bourdette, D., Pullicino, P., Scherokman, B and Whitham, R. (1992) Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). *Neurology* **42**, 859-863.

Goodkin, D.E., Hertsgaard, D. and Rudick, R.A. (1989) Exacerbation rates and adherence to disease type in a prospectively followed-up population with multiple sclerosis. Implications for clinical trials. *Archives of Neurology* **46**, 1107-1112.

Granger, C.V., Cotter, A.C., Hamilton, B.B., Fiedler, R.C. and Hens, M.M. (1990a) Functional assessment scales: a study of persons with multiple sclerosis. *Archives of Physical Medicine & Rehabilitation* **71**, 870-875.

Granger, C.V. and Gresham, G.E. (1990b) International Classification of Impairments, Disabilities, and Handicaps (ICIDH) as a conceptual basis for stroke outcome research. *Stroke* **21 Suppl II**, 66-7.

Griffin, L.D., Colchester, A.C.F., Roll, S.A. and Stundholm, C.S. (1996) Hierarchical segmentation satisfying constraints. In: Hancock, E.R., (Ed.) *Proceedings of the British Machine Vision Conference*, pp. 135-144. Sheffield: BMVA Press

Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, GJ, Plummer, D.L., Tofts, P.S., McDonald, W.I. and Miller, D.H. (1996) Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magnetic Resonance Imaging* **14** , 495-505.

Gronwall, D.M. (1977) Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual & Motor Skills* **44**, 367-373.

Hall, K.M., Hamilton, B.B., Gordon, W.A. and Zasler, N.D. (1993) Characteristics and comparisons of functional assessment indices: Disability Rating Scale, Functional Independence Measure, and Functional Assessment Measure. *Journal of Head Trauma Rehabilitation* **8**, 60-74.

Hamilton, B.B., Granger, C.V., Sherwin, F.S., Zielezny, M. and Teshman, J.S. (1987) A uniform national data system for medical rehabilitation. In: Fuhrer, M.J., (Ed.) *Rehabilitation outcomes: analysis and measurements*, pp. 137-147. Baltimore: Brookes

Hammond, S.R., McLeod, J.G., Millingen, K.S., Stewart-Wynne, E.G., English, D., Holland, J.T. and McCall, M.G. (1988) The epidemiology of multiple sclerosis in three Australian cities: Perth, Newcastle and Hobart. *Brain* **111**, 1-25.

Hartung, H.P., Archelos, J.J., Zielasek, J., Gold, R., Koltzenburg, Reiners, K.H. and Toyka, K.V. (1995) Circulating adhesion molecules and inflammatory mediators in demyelination: a review. *Neurology* **45**, S22-S32

Harwood, R.H., Gompertz, P. and Ebrahim, S. (1994) Handicap one year after a stroke: validity of a new scale. *Journal of Neurology, Neurosurgery & Psychiatry* **57**, 825-829.

Hauser, S.L., Dawson, D.M., Lehrich, J.R., Beal, M.F., Kevy, S.V., Propper, R.D., Mills, J.A. and Weiner, H.L. (1983) Intensive immunosuppression in progressive multiple sclerosis. A randomized, three-arm study of high-dose intravenous cyclophosphamide, plasma exchange, and ACTH. *New England Journal of Medicine* **308**, 173-180.

Herndon, R.M. and Goodkin, D.E. (1997) Multiple sclerosis and demyelination diseases. In: Herndon, R.M., (Ed.) *Handbook of neurologic rating scales*, pp. 107-123. New York: demos vermande

Hobart, J.C., Freeman, J.A. and Lamping, D.L. (1996a) Physician and patient-oriented outcomes in progressive neurological disease: which to measure?. *Current Opinion in Neurology* **9**, 441-444.

Hobart, J.C., Lamping, D.L., Freeman, J.A. and Thompson, A.J. (1996b) Measuring disability in multiple sclerosis: reliability of the functional independence measure. *European Journal of Neurology* **3 Suppl 2**, 123 (Abstract)

Hobart, J.C., Lamping, D.L., Freeman, J.A. and Thompson, A.J. (1996c) Reliability, validity and responsiveness of the Kurtzke expanded disability status scale (EDSS) in multiple sclerosis patients. *European Journal of Neurology* **3 Suppl 4**, 13 (Abstract)

Hobart, J.C., Lamping, D.L., Freeman, J.A. and Thompson, A.J. (1996d) The responsiveness of disability measures in multiple sclerosis. *European Journal of Neurology* **3 Supp 4**, 13 (Abstract)

Hobart, J.C., Lamping, D.L. and Thompson, A.J. (1996e) Evaluating neurological outcome measures: the bare essentials. *Journal of Neurology, Neurosurgery & Psychiatry* **60**, 127-130.

Holmes, J., Madgwick, T. and Bates, D. (1995) The cost of multiple sclerosis. *British Journal of Medical Economics* **8**, 181-193.

Holmes, W.C. and Shea, J.A. (1997) Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. *Quality of Life Research* **6**, 561-571.

Horsfield, M.A., Larsson, H.B., Jones, D.K. and Gass, A. (1997) Diffusion magnetic resonance imaging in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **64 Suppl 1**, S80-S84

Hughes, R. and Sharrack, B. (1998) Measuring disablement and quality of life in multiple sclerosis. *European Journal of Neurology* **5 Suppl 2**, S41-S42 (Abstract)

Hughes, R.A. and Sharrack, B. (1996) More immunotherapy for multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **61**, 239-241.

Hughes, R.A., Heywood, P. and Sharrack, B. (1995) Drug trials in neuropathy. *Bailliere's Clinical Neurology* **4**, 607-627.

Hunt, S.M., McKenna, S.P., McEwen, J., Williams, J. and Papp, E. (1981) The Nottingham Health Profile: subjective health status and medical consultations. *Social Science & Medicine - Part A, Medical Sociology* **15**, 221-229.

Hutchinson, J. and Hutchinson, M. (1995) The Functional Limitations Profile may be a valid, reliable and sensitive measure of disability in multiple sclerosis. *Journal of Neurology* **242**, 650-657.

International Federation of Multiple Sclerosis Societies (1985) *Minimal Record of Disability for Multiple Sclerosis*, New York: International Federation of Multiple Sclerosis Societies.

Jacobs, L.D., Cookfair, D.L., Rudick, R.A., Herndon, R.M., Richert, J.R., Salazar, A.M., Fischer, J.S., Goodkin, D.E., Granger, C.V., Simon, J.H., Alam, J.J., Bartoszak, D.M., Bourdette, D.N., Braiman, J., Brownscheidle, C.M., Coats, M.E., Cohan, S.L., Dougherty, DS, Kinkel, R.P., Mass, M.K., Munschauer, F.E., Priore, R.L., Pullicino, P.M., Scherokman, B.J. and Whitham, R.H. (1996) Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Annals of Neurology* **39**, 285-294.

Jonsson, B. and Henriksson, F. (1998) Economic burden of multiple sclerosis. *European Journal of Neurology* **5 Suppl 2**, S35-S37

Kaplan, R.M. and Anderson, J.P. (1988) A general health policy model: update and applications. *Health Services Research* **23**, 203-235.

Karnofsky, D.A. and Burchenal, J.H. (1949) Clinical evaluation of chemotherapeutic agents in cancer. In: McLeod, C.M., (Ed.) *Evaluation of chemotherapeutic agents*, pp. 191-205. New York: Coloubia University Press

Kazis, L.E., Anderson, J.J. and Meenan, R.F. (1989) Effect sizes for interpreting changes in health status. *Medical Care* **27**, S178-89.

Keith, R.A., Granger, C.V., Hamilton, B.B. and Sherwin, F.S. (1987a) The functional independence measure: a new tool for rehabilitation. *Advances in Clinical Rehabilitation* **1**, 6-18.

Kidd, D., Stewart, G., Baldry, J., Johnson, J., Rossiter, D., Petruckevitch, A. and Thompson, A.J. (1995) The Functional Independence Measure: a comparative validity and reliability study. *Disability & Rehabilitation* **17**, 10-14.

Koziol, J.A., Frutos, A., Sipe, J.C., Romine, J.S. and Beutler, E. (1996) A comparison of two neurologic scoring instruments for multiple sclerosis. *Journal of Neurology* **243**, 209-213.

Kretschmann, H. and Weinrich, W. (1992) *Cranial neuroimaging and clinical neuroanatomy*, Stuttgart: Georg Thieme Verlag.

Kurtzke, J.F. (1970) Clinical features of multiple sclerosis. In: Vinken, P.J. and Bruyn, G.W., (Eds.) *Handbook of clinical neurology*, pp. 161-216. Amsterdam: North-Holland

Kurtzke, J.F. (1983) Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33**, 1444-1452.

Kurtzke, J.F., Beebe, G.W. and Norman, J.E., Jr. (1979) Epidemiology of multiple sclerosis in U.S. veterans: 1. Race, sex, and geographic distribution. *Neurology* **29**, 1228-1235.

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **3**, 159-173.

LaRocca, N.G., Ritvo, P.G., Miller, D.M., Fisher, J.S., Andrews, H. and Paty, D.W. (1996) "Quality of life" assessment in multiple sclerosis clinical trials: current status and strategies for improving multiple sclerosis clinical design. In: Goodkin, D.E. and Rudick, R.A., (Eds.) *Multiple Sclerosis: advances in clinical trial design, treatment, and future perspectives*, pp. 145-160. New York: Springer

Lassmann, H. (1998) Pathology of multiple sclerosis. In: Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews, B. and Wekerle, H., (Eds.)

McAlpine's multiple sclerosis, 3 edn. pp. 323-358. London: Churchill Livingstone

Losseff, N.A., Wang, L., Lai, H.M., Yoo, D.S., Gawne-Cain, M.L., McDonald, W.I., Miller, D.H. and Thompson, A.J. (1996a) Progressive cerebral atrophy in multiple sclerosis. A serial MRI study. *Brain* **119**, 2009-2019.

Losseff, N.A., Webb, S.L., O'Riordan, J.I., Page, R., Wang, L., Barker, G.J., Tofts, P.S., McDonald, W.I., Miller, D.H. and Thompson, A.J. (1996b) Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain* **119**, 701-708.

Lublin, F.D. and Reingold, S.C. (1996) Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* **46**, 907-911.

Mahoney, F.I. and Barthel, D.W. (1965) Functional evaluation - the Barthel Index. *Maryland State Medical Journal* **14**, 61-65.

Martin, J., Melzer, H. and Elliot, D. (1988) *OPCS surveys of disability in Great Britain. I. The prevalence of disability among adults*, London: OPCS.

Mathiowetz, V., Weber, K., Kashman, N. and Volland, G. (1985) Adult norms for the nine hole peg test of finger dexterity. *Occupational Therapy Journal of Research* **5**, 24-38.

Matthews, B. (1998) Symptoms and signs of multiple sclerosis. In: Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews, B. and Wekerle, H., (Eds.) *McAlpine's multiple sclerosis*, 3 edn. pp. 145-190. London: Churchill Livingstone

Matthews, W.B. (1991) Multiple sclerosis and other demyelinationg diseases: clinical features. In: Swash, M. and Oxbury, J., (Eds.) *Clinical neurology*, 1 edn. pp. 1106-1115. London: Churchill Livingstone

McAlpine D (1972) *Multiple sclerosis: a reappraisal*, Edinburgh: Chirchill Livingstone.

McDonald, I. (1998) Pathophysiology of multiple sclerosis. In: Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews, B. and Wekerle, H., (Eds.) *McAlpine's multiple sclerosis*, 3 edn. pp. 359-378. London: Churchill Livingstone

McDonald, W.I., Miller, D.H. and Thompson, A.J. (1994) Are magnetic resonance findings predictive of clinical outcome in therapeutic trials in multiple sclerosis? The dilemma of interferon-beta. *Annals of Neurology* **36**, 14-18.

McGowan, J.C., Filippi, M., Campi, A. and Grossman, R.I. (1997) Magnetisation transfer imaging: theory and application to multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **64 Suppl 1**, S66-S69

Mclever, J.P. and Carmines, E.G. (1981) *Unidimentional scaling*, Newbury Park, California: Sage.

Midgard, R., Riise, T. and Nyland, H. (1996) Impairment, disability and handicap in multiple sclerosis: a cross-sectional study in an incident cohort in More and Romsdal County, Norway. *Journal of Neurology* **243**, 337-344.

Miller, D.H., Albert, P.S., Barkhof, F., Francis, G., Frank, J.A., Hodgkinson, S., Lublin, F.D., Paty, D.W., Reingold, S.C. and Simon, J. (1996) Guidelines for the use of magnetic resonance techniques in monitoring the treatment of multiple sclerosis. US National MS Society Task Force. *Annals of Neurology* **39**, 6-16.

Miller, D.H., Grossman, R.I., Reingold, S.C. and McFarland, H.F. (1998) The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain* **121**, 3-24.

Miller, E. (1984) Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology* **23**, 53-57.

Moll, C., Mourre, C., Lazdunski, M. and Ulrich, J. (1991) Increase of sodium channels in demyelinated lesions of multiple sclerosis. *Brain Research* **556**, 311-316.

Moor, G.R.W. (1998) Neuropathology and pathophysiology of the multiple sclerosis lesion. In: Paty, D.W. and Ebers, G., (Eds.) *Multiple sclerosis*, 1 edn. pp. 257-327. Philadelphia: F.A. Davis Company

Morrissey, S.P., Miller, D.H., Kendall, B.E., Kingsley, D.P., Kelly, MA, Francis, D.A., MacManus, D.G. and McDonald, W.I. (1993) The significance of brain magnetic resonance imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple sclerosis. A 5-year follow-up study. *Brain* **116**, 135-146.

Mumford, C.J. and Compston, A. (1993) Problems with rating scales for multiple sclerosis: a novel approach--the CAMBS score. *Journal of Neurology* **240**, 209-215.

Mumford, C.J., Wood, N.W., Kellar-Wood, H., Thorpe, J.W., Miller, DH and Compston, D.A. (1994) The British Isles survey of multiple sclerosis in twins. *Neurology* **44**, 11-15.

Myers, L.W., Leake, B.D. and Ellison, G.W. (1992) Use of survival analysis to describe the course of multiple sclerosis. *Annals of Neurology* **32**, 259-259 (Abstract)

Nathadwarawala, K.M., Nicklin, J. and Wiles, C.M. (1992) A timed test of swallowing capacity for neurological patients. *Journal of Neurology, Neurosurgery & Psychiatry* **55**, 822-825.

Nelson, H.E. and Willison, J. (1992) *Restandaraisation of the NART against the WAIS-R*, Windsor UK: National Federation for Educational Research.

Norman, G.R., Stratford.P. and Regehr, G. (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* **50**, 869-879.

Norman, G.R. and Streiner, D.L. (1993a) Comparing two groups, the *t*-test. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 58-63. St. Lois, Missouri: Mosby

Norman, G.R. and Streiner, D.L. (1993b) Elements of statistical inference. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 38-52. St. Lois, Missouri: Mosby

Norman, G.R. and Streiner, D.L. (1993c) Multiple regression. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 108-118. St. Lois, Missouri: Mosby

Norman, G.R. and Streiner, D.L. (1993d) Principal components and factor analysis. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 129-142. St. Lois, Missouri: Mosby

Norman, G.R. and Streiner, D.L. (1993e) Simple regression and correlation. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 100-107. St. Lois, Missouri: Mosby

Norman, G.R. and Streiner, D.L. (1993f) Tests of significance for ranked data. In: Norman, G.R. and Streiner, D.L., (Eds.) *Biostatistics*, pp. 170-175. St. Lois, Missouri: Mosby

Noseworthy, J.H. (1994) Clinical scoring methods for multiple sclerosis. *Annals of Neurology* **36 Suppl**, S80-S85

Noseworthy, J.H., Vandervoort, M.K., Wong, C.J. and Ebers, G.C. (1990) Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. *Neurology* **40**, 971-975.

Nunnally, J.C. (1978) *Psychometric theory*, 2nd edn. New York: McGraw-Hill.

O'Brien, B. (1987) *Multiple sclerosis*, London: Office of Health Economics.

O'Riordan, J.I., Thompson, A.J., Kingsley, D.P., MacManus, D.G., Kendall, B.E., Rudge, McDonald, W.I. and Miller, D.H. (1997) The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain* **121**, 495-503.

Olerup, O. and Hillert, J. (1991) HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* **38**, 1-15.

Patrick, D.L. and Deyo, R.A. (1989) Generic and disease-specific measures in assessing health status and quality of life. *Medical Care* **27**, S217-32.

Patrick, D.L., Sittampalam, Y., Somerville, S.M., Carter, W.B. and Bergner, M. (1985) A cross-cultural comparison of health status values. *American Journal of Public Health* **75**, 1402-1407.

Paty, D.W. (1999) Results of the 3 year double-blind, placebo-controlled study of interferon beta 1a (Rebif) in secondary progressive MS. *Journal of Neurology* **246 Suppl 1**, I/15c(Abstract)

Paty, D.W. (1988) Magnetic resonance imaging in the assessment of disease activity in multiple sclerosis. *Canadian Journal of Neurological Sciences* **15**, 266-272.

Paty, D.W. and Li, D.K. (1993) Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* **43**, 662-667.

Paty, D.W., Li, D.K., Oger, J.J., Kastrukoff, L., Koopmans, R., Tanton, E. and Zhao, G.J. (1994) Magnetic resonance imaging in the evaluation of clinical trials in multiple sclerosis. *Annals of Neurology* **36 Suppl**, S95-S96

Paty, D.W. and Moor, G.R.W. (1997) Magnetic resonance imagings as living pathology in multiple sclerosis. In: Paty, D.W. and Ebers, G.C., (Eds.) *Multiple sclerosis*, pp. 328-369. Phladephia: F.A. Davis Company

Peto, V., Jenkinson, C., Fitzpatrick, R. and Greenhall, R. (1995) The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Quality of Life Research* **4**, 241-248.

Phillips, V.L. (1995) Community care for severely disabled people on low incomes. *British Medical Journal* **311**, 1121-1123.

Poser, C.M., Paty, D.W., Scheinberg, L., McDonald, W.I., Davis, F.A., Ebers, G.C., Johnson, K.P., Sibley, W.A., Silberberg, D.H. and Tourtellotte, W.W. (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Annals of Neurology* **13**, 227-231.

Poser, S., Kurtzke, J.F., Poser, W. and Schlaf, G. (1989) Survival in multiple sclerosis. *Journal of Clinical Epidemiology* **42**, 159-168.

Poser, S., Wikstrom, J. and Bauer, H.J. (1979) Clinical data and the identification of special forms of multiple sclerosis in 1271 cases studied with a standardized documentation system. *Journal of the Neurological Sciences* **40**, 159-168.

PRISMS Study Group (1998) Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. *Lancet* **352**, 1498-1504.

Raine, C.S. (1991) Demyelinating diseases. In: Davis,R.L. and Robertson, D.M., (Eds.) *Textbook of neuropathology*, 2 edn. pp. 535-620. Baltimore: Willams & Wilkins

Rao, S.M., Leo, G.J., Haughton, V.M., St and Bernardin, L. (1989) Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. *Neurology* **39**, 161-166.

Raven, J.C. (1994) *Standard progressive matrices*, Oxford: Oxford Psychologists Press.

Rodriguez, M., Siva, A., Ward, J., Stolp-Smith, K., O'Brien, P. and Kurland, L. (1994) Impairment, disability, and handicap in multiple sclerosis: a population-based study in Olmsted County, Minnesota. *Neurology* **44**, 28-33.

Ross, M. (1989) Relation of implicit theories to the construction of personal histories. *Psychological Review* **96**, 341-347.

Rothwell, P.M., McDowell, Z., Wong, C.K. and Dorman, P.J. (1997) Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *British Medical Journal* **314**, 1580-1583.

Rudick, R., Antel, J., Confavreux, C., Cutter, G., Ellison, G., Fischer, J., Lublin, F., Miller, A., Petkau, J., Rao, S., Reingold, S., Syndulko, K., Thompson, A., Wallenberg, J., Weinshenker, B. and Willoughby, E. (1996a) Clinical outcomes assessment in multiple sclerosis. *Annals of Neurology* **40**, 469-479.

Rudick, R., Antel, J., Confavreux, C., Cutter, G., Ellison, G., Fischer, J., Lublin, F., Miller, A., Petkau, J., Rao, S., Reingold, S., Syndulko, K., Thompson, A., Wallenberg, J., Weinshenker, B. and Willoughby, E. (1996b) Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. *Annals of Neurology* **42**, 379-382.

Runmarker, B. and Andersen, O. (1993) Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain* **116**, 117-134.

Rust, J. and Golombok, S. (1985) The Golombok-Rust Inventory of Sexual Satisfaction (GRISS). *British Journal of Clinical Psychology* **24**, 63-64.

Sadovnick, A.D., Armstrong, H., Rice, G.P., Bulman, D., Hashimoto, Paty, D.W., Hashimoto, S.A., Warren, S., Hader, W. and Murray, T.J. (1993) A population-

based study of multiple sclerosis in twins: update. *Annals of Neurology* **33**, 281-285.

Sadovnick, A.D., Dyment, D. and Ebers, G.C. (1996) Genetic epidemiology of multiple sclerosis. *Epidemiologic Reviews* **19**, 99-106.

Schipper, H., Clinch, J. and Powell, V. (1990) Definitions and conceptual issues. In: Spiker, B., (Ed.) *Quality of life assessments in clinical trials*, pp. 5-24. New York: Raven Press

Seedhouse, D. (1994) *Ethics: the heart of health care*, Chichester: John Wiley & Sons Ltd.

Sharrack, B. and Hughes, R.A. (1996) Clinical scales for multiple sclerosis. *Journal of the Neurological Sciences* **135**, 1-9.

Sharrack, B. and Hughes, R.A. (1997) Reliability of distance estimation by doctors and patients: cross sectional study. *British Medical Journal* **315**, 1652-1654.

Sharrack, B. and Hughes, R.A. (1999a) Scale development and the Guy's Neurological Disability Scale. *Journal of Neurology* **246**, 226.

Sharrack, B. and Hughes, R.A. (1999b) The Guy's Neurological Disability Scale: a new disability scale for multiple sclerosis. *Multiple Sclerosis* **5**, 223-233.

Sharrack, B., Hughes, R.A., Soudain, S. and Dunn, G. (1999c) The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* **122**, 141-159.

Shepherd, D.I. (1979) Clinical features of multiple sclerosis in north-east Scotland. *Acta Neurologica Scandinavica* **60**, 218-230.

Shibasaki, H., McDonald, W.I. and Kuroiwa, Y. (1981) Racial modification of clinical picture of multiple sclerosis: comparison between British and Japanese patients. *Journal of the Neurological Sciences* **49**, 253-271.

Shrout, P.E. and Fleiss, J.L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* **86**, 420-428.

Simon, J.H., Jacobs, L.D., Campion, M., Wende, K., Simonian, N., Cookfair, D.L., Rudick, R.A., Herndon, R.M., Richert, JR, Salazar, A.M., Alam, J.J., Fischer, J.S., Goodkin, D.E., Granger, C.V., Lajaunie, Martens-Davidson, A.L., Meyer, M., Sheeder, J., Choi, K., Scherzinger, A.L., Bartoszak, D.M., Bourdette, D.N., Braiman, J., Brownscheidle, C.M. and Whitham, R.H. (1998) Magnetic resonance studies of intramuscular interferon beta-1a for relapsing multiple sclerosis. *Annals of Neurology* **43**, 79-87.

Sinclair, D.J. and Ingram, C.G. (1980) Controlled trial of supervised exercise training in chronic bronchitis. *British Medical Journal* **280**, 519-521.

Sipe, J.C., Knobler, R.L., Braheny, S.L., Rice, G.P., Panitch, H.S. and Oldstone, M.B. (1984) A neurologic rating scale (NRS) for use in multiple sclerosis. *Neurology* **34**, 1368-1372.

Smith, D. (1993) Review of quality of life. In: Chadwick, D.W., Baker, G.A. and Jacoby, A., (Eds.) *Quality of life and quality of care in epilepsy: update 1993*, pp. 5-13. London: Royal Society of Medicine Services Limited

Steadman, P. (1992) Applying for disability living allowance. *British Medical Journal* **305**, 893

Steadman, P. (1993) The new disability living allowance. *Archives of Disease in Childhood* **68**, 73-74.

Stone, L.A., Smith, M.E., Albert, P.S., Bash, C.N., Maloni, H., Frank, JA and McFarland, H.F. (1995) Blood-brain barrier disruption on contrast-enhanced MRI in patients with mild relapsing-remitting multiple sclerosis: relationship to course, gender, and age. *Neurology* **45**, 1122-1126.

Streiner, D.L. and Norman, G.R. (1995a) Biases in responding. In: Streiner, D.L. and Norman, G.R., (Eds.) *Health measurement scales*, Second edn. pp. 69-84. New York: Oxford University Press

Streiner, D.L. and Norman, G.R. (1995b) Reliability. In: Streiner, D.L. and Norman, G.R., (Eds.) *Health measurement scales*, Second edn. pp. 104-127. New York: Oxford University Press

Streiner, D.L. and Norman, G.R. (1995c) Selecting the items. In: Streiner, D.L. and Norman, G.R., (Eds.) *Health measurement scales*, Second edn. pp. 54-68. New York: Oxford University Press

Swingler, R.J. and Compston, D.A. (1992) The morbidity of multiple sclerosis. *Quarterly Journal of Medicine* **83**, 325-337.

Testa, M.A. and Simonson, D.C. (1996) Assessment of quality-of-life outcomes. *New England Journal of Medicine* **334**, 835-840.

The Canadian Burden of Illness Study Group (1998) Burden of illness of multiple sclerosis: Part I: Cost of illness. *Canadian Journal of Neurological Sciences* **25**, 23-30.

The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group (1995) Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* **45**, 1277-1285.

Thompson, A.J. (1996a) Multiple sclerosis: symptomatic treatment. *Journal of Neurology* **243**, 559-565.

Thompson, A.J. and Hobart, J.C. (1996b) Multiple sclerosis: assessment of disability and disability scales. *Journal of Neurology* **245**, 189-196.

Thompson, A.J., Kermode, A.G., MacManus, D.G., Kendall, B.E., Kingsley, D.P., Moseley, I.F. and McDonald, W.I. (1990) Patterns of disease activity in

multiple sclerosis: clinical and magnetic resonance imaging study. *British Medical Journal* **300**, 631-634.

Thompson, A.J. and Noseworthy, J.H. (1996) New treatments for multiple sclerosis: a clinical perspective. *Current Opinion in Neurology* **9**, 187-198.

Thorpe, J.W., Halpin, S.F., MacManus, D.G., Barker, G.J., Kendall, BE and Miller, D.H. (1994) A comparison between fast and conventional spin-echo in the detection of multiple sclerosis lesions. *Neuroradiology* **36**, 388-392.

Thuriaux, M.C. (1995) The ICIDH: evolution, status, and prospects The ICIDH: evolution, status, and prospects. *Disability & Rehabilitation* **17**, 112-118.

Trapp, B.D., Peterson, J., Ransohoff, R.M., Rudick, R., Mork, S. and Bo, L. (1996) Axonal transection in the lesions of multiple sclerosis. *New England Journal of Medicine* **338**, 278-285.

Truyen, L., van Waesberghe, J.H., van, Walderveen, M.A., van Oosten, B.W., Polman, C.H., Hommes, O.R., Ader, H.J. and Barkhof, F. (1996) Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. *Neurology* **47**, 1469-1476.

Utz, U. and McFarland, H.F. (1994) The role of T cells in multiple sclerosis: implications for therapies targeting the T cell receptor. *Journal of Neuropathology & Experimental Neurology* **53**, 351-358.

Valleroy, M.L. and Kraft, G.H. (1984) Sexual dysfunction in multiple sclerosis. *Archives of Physical Medicine & Rehabilitation* **65**, 125-128.

van der Putten, J.J., Hobart, J.C., Freeman, J.A. and Thompson, A.J. (1999) Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery & Psychiatry* **66**, 480-484.

van Waesberghe, J.H., van, Walderveen, M.A., Castelijns, J.A., Scheltens, P., Lycklama, Polman, CH and Barkhof, F. (1997) Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR. *American Journal of Neuroradiology* **19**, 675-683.

van Walderveen, M.A., Barkhof, F., Hommes, O.R., Polman, C.H., Tobi, H., Frequin, S.T. and Valk, J. (1995) Correlating MRI and clinical disease activity in multiple sclerosis: relevance of hypointense lesions on short-TR/short-TE (T1-weighted) spin-echo images. *Neurology* **45**, 1684-1690.

van Walderveen, M.A., Kamphorst, W., Scheltens, van Waesberghe, J.H., Ravid, R., Valk, J., Polman, C.H. and Barkhof, F. (1996) Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. *Neurology* **50**, 1282-1288.

Vickrey, B.G., Hays, R.D., Engel, J., Jr., Spritzer, K., Rogers, W.H., Rausch, R., Graber, J. and Brook, R.H. (1995a) Outcome assessment for epilepsy surgery: the impact of measuring health-related quality of life. *Annals of Neurology* **37**, 158-166.

Vickrey, B.G., Hays, R.D., Harooni, R., Myers, L.W. and Ellison, G.W. (1995b) A health-related quality of life measure for multiple sclerosis. *Quality of Life Research* **4**, 187-206.

Vickrey, B.G., Hays, R.D. and Spritzer, K. (1993) Methodologic issues in quality of life assessment for epilepsy surgery. In: Chadwick, D.W., Baker, G.A. and Jacoby, A., (Eds.) *Quality of life and quality of care in epilepsy: update 1993*, pp. 27-37. London: Royal sciety of Medicine Services Limited

Wade, D.T. (1995a) Pathology, impairment, disability, handicap: a useful model. In: Wade, D.T., (Ed.) *Measurement in neurological rehabilitation*, pp. 3-14. New York: Oxford University Press

Wade, D.T. (1995b) Personal physical disability. In: Wade, D.T., (Ed.) *Measurement in neurological rehabilitation*, pp. 70-82. New York: Oxford University Press

Wade, D.T. (1996) Epidemiology of disabling neurological disease: how and why does disability occur?. *Journal of Neurology, Neurosurgery & Psychiatry* **61**, 242-249.

Wechsler, D. (1981) *Wechsler adult intelligence scale - revised*, San Antonio, Texas: Psychological Corporation.

Weinshenker, B.G. (1994) Natural history of multiple sclerosis. *Annals of Neurology* **36 Suppl**, S6-11.

Weinshenker, B.G., Bass, B., Rice, G.P., Noseworthy, J., Carriere, Baskerville, J. and Ebers, G.C. (1989a) The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain* **112**, 1419-1428.

Weinshenker, B.G., Bass, B., Rice, G.P., Noseworthy, J., Carriere, Baskerville, J. and Ebers, G.C. (1989b) The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. *Brain* **112**, 133-146.

Whetten-Goldstein, K., Sloan, F.A., Goldstein, L.B. and Kulas, E.D. (1998) A comprehensive assessment of the cost of multiple sclerosis in the United States. *Multiple Sclerosis* **4**, 419-425.

Whitaker, J.N., McFarland, H.F., Rudge, P. and Reingold, S.C. (1995) Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple Sclerosis* **1**, 37-47.

Willison, J.R., Thomas, D.J., du, B.G., Marshall, J., Paul, EA, Pearson, T.C., Russell, R.W., Symon, L. and Wetherley-Mein, G. (1980) Effect of high haematocrit on alertness. *Lancet* **1**, 846-848.

Willoughby, E.W. and Paty, D.W. (1988) Scales for rating impairment in multiple sclerosis: a critique. *Neurology* **38** , 1793-1798.

World Health Organisation (1947) *The constitution of the World Health Organisation*, Geneva: World Health Organisation.

World Health Organisation (1980) *International Classification of Impairments, Disabilities, and Handicaps*, Geneva: World Health Organisation.

World Health Organisation (1993) *International Statistical Classification of Disease and Related Health Problems (CD-10)*, Geneva: World Health Organisation.

World Health Organisation (1997) *ICIDH-2: International Classification of Impairments, Activities, and Participation*, Beta-1 draft edn. Geneva: World Health Organaisation.